GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 12:59:04 ; Search time 5686.68 Seconds
(without alignments)
5067.252 Million cell updates/sec

Title: US-09-997-610-3
Perfect score: 1377
Sequence: 1 athgtngtnathccngtnyt.....gyytntayytncaytgyatg 1377

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	921	66.9	145880	9	HS302D9	Z82198 Human DNA s
2	497.8	36.2	154090	9	AC025577	AC025577 Homo sapi
3	497.8	36.2	186660	2	AC026107	AC026107 Homo sapi
c 4	493.6	35.8	124518	9	AL138962	AL138962 Human DNA
5	493	35.8	123070	2	AC008799	AC008799 Homo sapi
c 6	493	35.8	189768	2	AC044889	AC044889 Homo sapi
7	490.4	35.6	198545	2	AC017063	AC017063 Homo sapi
c 8	488.8	35.5	161575	2	AC025233	AC025233 Homo sapi
9	488.8	35.5	163674	2	AC087500	AC087500 Homo sapi
c 10	488.8	35.5	175639	2	AC015727	AC015727 Homo sapi
c 11	488.6	35.5	181842	2	AL391823	AL391823 Homo sapi
12	488.2	35.5	77405	2	AL353634	AL353634 Homo sapi
c 13	488.2	35.5	124531	9	HSJ633H17	AL049710 Human DNA
14	487.6	35.4	40714	9	HSU212C1	Z69722 Human DNA s
c 15	487.6	35.4	159712	9	AP004219	AP004219 Homo sapi
c 16	487	35.4	125295	2	AL672061	AL672061 Homo sapi
c 17	487	35.4	146743	2	AC093588	AC093588 Homo sapi
c 18	487	35.4	152544	9	CNS05TEJ	AL359232 Human chr
c 19	487	35.4	166679	9	AC079899	AC079899 Homo sapi
c 20	487	35.4	174662	2	AC026036	AC026036 Homo sapi
c 21	487	35.4	176426	9	AC007370	AC007370 Homo sapi
c 22	487	35.4	203726	2	AC011882	AC011882 Homo sapi
23	486	35.3	139005	2	AF286112	AF286112 Homo sapi
c 24	486	35.3	156458	2	AC025647	AC025647 Homo sapi
c 25	486	35.3	165901	9	AP003474	AP003474 Homo sapi
c 26	486	35.3	171427	9	AC021590	AC021590 Homo sapi
c 27	485.8	35.3	99084	2	AC026420	AC026420 Homo sapi
c 28	485.8	35.3	133790	9	AC010280	AC010280 Homo sapi
c 29	485.8	35.3	150332	9	AC004921	AC004921 Homo sapi
30	485.8	35.3	174293	9	AC008816	AC008816 Homo sapi
31	485.8	35.3	203043	9	AC017093	AC017093 Homo sapi
32	485.4	35.3	131215	9	AC079614	AC079614 Homo sapi
33	485	35.2	107885	9	AC006389	AC006389 Homo sapi
34	484.6	35.2	173480	9	CNS00M8T	AL079343 Human chr
c 35	483	35.1	164370	9	AC097460	AC097460 Homo sapi
36	482.6	35.0	112515	9	AL139216	AL139216 Human DNA
37	482.6	35.0	142224	2	AC096535	AC096535 Homo sapi
38	482.6	35.0	178535	2	AL354978	AL354978 Homo sapi
c 39	482	35.0	168502	9	AC091005	AC091005 Homo sapi
40	481.6	35.0	38235	9	AC004559	AC004559 Homo sapi
41	481.4	35.0	234622	2	AC110089	AC110089 Homo sapi
c 42	479.8	34.8	187883	9	AC008268	AC008268 Homo sapi
43	479.6	34.8	167891	9	AC013439	AC013439 Homo sapi
c 44	479.2	34.8	73390	9	AC012038	AC012038 Homo sapi
45	479.2	34.8	94730	9	AP000230	AP000230 Homo sapi

ALIGNMENTS

RESULT 1
HS302D9
LOCUS HS302D9 145880 bp DNA linear PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains GSSs, complete sequence.
ACCESSION Z82198
VERSION Z82198.2 GI:6572207
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 145880)
AUTHORS Bridgeman,A.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Dec 13, 1999 this sequence version replaced gi:3164067.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22
RP1-302D9 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP1-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence.

FEATURES Location/Qualifiers
source 1. .145880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP1-302D9"
/clone_lib="RPCI-1"
repeat_region 188. .245
/note="MER3 repeat: matches 144. .209 of consensus"
repeat_region 246. .571
/note="AluSx repeat: matches 1. .312 of consensus"
repeat_region 572. .759
/note="MER3 repeat: matches 1. .144 of consensus"
repeat_region 783. .933
/note="MER5A repeat: matches 26. .187 of consensus"
repeat_region 1033. .1336
/note="AluSp repeat: matches 1. .299 of consensus"
repeat_region 1450. .1583
/note="MIR repeat: matches 24. .160 of consensus"
repeat_region 1687. .1752
/note="L2 repeat: matches 2593. .2661 of consensus"
repeat_region 2350. .2660
/note="AluSc repeat: matches 3. .309 of consensus"
repeat_region 2684. .2981
/note="AluSq repeat: matches 2. .300 of consensus"
repeat_region 3323. .3343
/note="MLT1E repeat: matches 116. .136 of consensus"
repeat_region 3344. .3652
/note="AluY repeat: matches 1. .309 of consensus"
repeat_region 3653. .3928
/note="MLT1E repeat: matches 136. .359 of consensus"
repeat_region 3929. .4278
/note="THE1B repeat: matches 3. .364 of consensus"
repeat_region 4279. .4485
/note="MLT1E repeat: matches 359. .568 of consensus"
repeat_region 5073. .5176
/note="52 copies 2 mer ct 78 conserved"
repeat_region 5181. .5491
/note="AluJb repeat: matches 1. .311 of consensus"
repeat_region 6369. .6485
/note="L2 repeat: matches 2579. .2705 of consensus"
repeat_region 6647. .6685
/note="MADE1 repeat: matches 1. .23 of consensus"
repeat_region 6686. .6987
/note="AluSx repeat: matches 1. .302 of consensus"
repeat_region 6988. .7036
/note="MADE1 repeat: matches 23. .77 of consensus"
repeat_region 7482. .7754
/note="AluJb repeat: matches 9. .290 of consensus"
repeat_region 7775. .8060
/note="AluJo repeat: matches 1. .295 of consensus"
repeat_region 8414. .8551
/note="L2 repeat: matches 2553. .2706 of consensus"
repeat_region 8914. .9030
/note="MIR repeat: matches 147. .262 of consensus"
repeat_region 9110. .9280
/note="MIR repeat: matches 91. .262 of consensus"
repeat_region 9283. .9412
/note="MIR repeat: matches 15. .144 of consensus"
repeat_region 9521. .9679
/note="FAM repeat: matches 3. .161 of consensus"
repeat_region 9820. .10225
/note="MSTB repeat: matches 2. .425 of consensus"
misc_feature complement(10179. .10678)
/note="match: GSS: Em:B56592"
misc_feature complement(10204. .10728)
/note="match: GSS: Em:AQ701486"
misc_feature complement(10249. .10706)
/note="match: GSS: Em:AQ225495"
repeat_region 10312. .10383
/note="MIR repeat: matches 79. .150 of consensus"
misc_feature 10718. .11310
/note="match: GSS: Em:B14024"
misc_feature 10784. .11201
/note="match: GSS: Em:B43656"
repeat_region 11838. .11946
/note="MIR repeat: matches 20. .137 of consensus"
repeat_region 12174. .12445
/note="L2 repeat: matches 1988. .2275 of consensus"
repeat_region 12444. .12642
/note="MIR repeat: matches 63. .241 of consensus"
misc_feature 13017. .13369
/note="match: STS: Em:G49301"
repeat_region 13331. .13397
/note="MIR repeat: matches 174. .244 of consensus"
repeat_region 13398. .13698
/note="AluSp repeat: matches 1. .302 of consensus"
repeat_region 13699. .13810
/note="MIR repeat: matches 76. .174 of consensus"
repeat_region 13806. .13919
/note="MIR repeat: matches 77. .189 of consensus"
repeat_region 13945. .14060
/note="MIR repeat: matches 24. .142 of consensus"
repeat_region 14061. .14367
/note="AluY repeat: matches 1. .301 of consensus"
repeat_region 14368. .14452
/note="MIR repeat: matches 141. .225 of consensus"
repeat_region 14589. .14679
/note="MIR repeat: matches 173. .262 of consensus"
misc_feature 14597. .15201
/note="match: GSS: Em:AQ553482"
misc_feature 14616. .15060
/note="match: GSS: Em:AQ370601"
repeat_region 14868. .15040
/note="MIR repeat: matches 49. .233 of consensus"
repeat_region 15071. .15188
/note="L2 repeat: matches 2112. .2239 of consensus"
repeat_region 15304. .15399
/note="MLT1B repeat: matches 1. .99 of consensus"
repeat_region 15490. .15662
/note="AluSg1 repeat: matches 2. .114 of consensus"
repeat_region 15669. .15727
/note="MLT1B repeat: matches 119. .178 of consensus"
repeat_region 15728. .16027
/note="AluSc repeat: matches 1. .299 of consensus"
repeat_region 16028. .16245
/note="MLT1B repeat: matches 178. .390 of consensus"
repeat_region 16546. .16854
/note="AluY repeat: matches 1. .300 of consensus"
repeat_region 18296. .18323
/note="MSTA repeat: matches 2. .29 of consensus"
repeat_region 18324. .18392

```
                     /note="MER66-internal repeat: matches 4919. .4993 of
                     consensus"
     repeat_region   18393. .18712
                     /note="AluJb repeat: matches 1. .311 of consensus"
     repeat_region   18713. .19133
                     /note="MER66-internal repeat: matches 4548. .4919 of
                     consensus"
     misc_feature    complement(18872. .19230)
                     /note="match: GSS: Em:AQ005063"
     misc_feature    19251. .19719
                     /note="match: GSS: Em:B14179"
     repeat_region   19537. .20290
                     /note="HERVFH21 repeat: matches 4657. .5784 of consensus"
     repeat_region   20317. .20382
                     /note="33 copies 2 mer ta 68 conserved"
     repeat_region   20513. .20666
                     /note="77 copies 2 mer tt 70 conserved"
     repeat_region   20682. .21008
                     /note="AluSg1 repeat: matches 1. .306 of consensus"
     repeat_region   21239. .21553
                     /note="HUERS-P3 repeat: matches 4410. .4713 of consensus"
     repeat_region   21882. .22254
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   22302. .22537
                     /note="MER66-internal repeat: matches 2186. .2417 of
                     consensus"
     repeat_region   22538. .22850
                     /note="AluSp repeat: matches 1. .313 of consensus"
     repeat_region   22851. .23801
                     /note="MER66-internal repeat: matches 1210. .2186 of
                     consensus"
     repeat_region   23905. .23989
                     /note="MER66-internal repeat: matches 3017. .3102 of

  Query Match             66.9%;  Score 921;  DB 9;  Length 145880;
  Best Local Similarity   58.2%;  Pred. No. 2e-211;
  Matches  750;  Conservative  300;  Mismatches  239;  Indels    0;  Gaps    0;

Qy      89 tnggnccnccnggngcnccnggnytnccncartayacnggngarathwsngaratgacna 148
           | |     || || || || || :| || ||:||:|| || ||:||::: ||:|||||  |
Db   37254 TGGAATGTCCTGGTGCCCCAGGTTTACCACAATATACAGGAGAAATAAGTGAAATGACAA 37313

Qy     149 artgyccntgyccngayathgarmgnwsngcnttyacngtnaarytnwsnggnaarytnc 208
           |:||:|| ||:|| ||:||:||::| :: || ||:|| || ||:::| :: || ||::| |
Db   37314 AATGCCCCTGTCCTGATATAGAAAGGTCAGCCTTTACTGTGAAGCTCAGTGGAAAACTTC 37373

Qy     209 cnytnccnttyaarccnathathttyacnggngtnytntayaaygcncarmgngayytna 268
           | :| || ||:||:|| ||:||:||:|| || || :| ||:||:|| ||::| ||::| |
Db   37374 CTCTTCCTTTCAAGCCCATCATCTTCACAGGGGTCCTGTACAATGCCCAGAGGGATTTAA 37433

Qy     269 argargcnatgggngtnttygcntgymgngtnccnggnaaytaytaywsnwsnttygayg 328
           |:||:|| ||||| || ||:|| ||::| || || || ||:||:||::: :: ||:||:|
Db   37434 AGGAGGCCATGGGAGTCTTTGCTTGCAGGGTGCCTGGGAATTACTACTCCAGCTTTGATG 37493

Qy     329 tngarytncaycaytgyaargtnaayathtggytnatgmgnaarcarathytngcnaaya 388
           | ||::| ||:||:||:||:|| ||:||:|||:| |||:| ||:||:||::| || ||:|
Db   37494 TTGAGCTGCATCATTGCAAGGTGAATATTTGGCTAATGAGGAAGCAAATTTTGGCTAATA 37553

Qy     389 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnt 448
           |:||:||:||::: ||:||:||::: ||:||:||:|| || ||||| :| :| ||:|| |
Db   37554 AGGAAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGGCAT 37613

Qy     449 tywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtna 508
           |::: ||:||::| ||:|| ||:||:||::: :: ||:||::| ||:|| ||:||:|| |
Db   37614 TCAGTTTCATAAGGGAGGCAGAGCATAAGAGTTCAGAAAATTTGCACCCTGACAATGTGA 37673

Qy     509 thaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyatht 568
           |:||:||:||:||:|| ||::: ||:|| ||:||:||::| || || ||:||:||:||:|
Db   37674 TAAAAAAGAAAAACCCATTTTCTGAGGGGAAATTCAAGCTGGCTGCAGAAATTTGCATAT 37733

Qy     569 gyaaygargarytnaaygtnaaycccncargayaayggngaraayathwsntggacntgyc 628
           |:||:||:||::| ||:|| ||:|| ||:||:||:|| ||:||:||::: ||||| ||:|
Db   37734 GTAATGAGGAGCTGAATGTTAATCCTCAAGACAATGGGGAAAATATCTCCTGGACATGTC 37793

Qy     629 armgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttyt 688
           |::| :: :: ||:||::: ||:||::: :| || |||:| || :| :| ||:|||||:|
Db   37794 AGAGGTCTTCACAGCAGTCCATCAAATCACTGGCCTGGAGGCCTAGGAGAAAATGGTTTT 37853

Qy     689 gyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtnc 748
           |:|| || || || || :: :| ||:||:|| ||:|| :| ||::| || || ||:|| |
Db   37854 GTGGGACAGGCCCAGGGTCCCTGTGCTGTGTGCAGCCTAGAGACTTGGTGCCCTGTGTCC 37913

Qy     749 cngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsng 808
           | || ||::: || || || :: ||:|| || :: || ||:|| |||||::| || :: |
Db   37914 CAGTTAATTCAGCTGTGGCTTCAGAGGGTGCAAGCCCCAAGCCTTGGCAGCTTCCAAGTG 37973

Qy     809 gngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnt 868
           | || ||:|| || || || ||:||::: :| ||:||:|| |||||:|| || ||::| |
Db   37974 GTGTTGAGCCTGTGGGTGCAAAGAAGTCAAGAATTGAGGTTTGGGAACCTCCAATCAGAT 38033

Qy     869 tycaraarathtayggnaayccntggatgccnmgncaraarttygcngtnggngtnggnw 928
           |:||:||:||:||:|| ||:|| |||||||| :| ||:||:||:|| || || || || :
Db   38034 TTCAGAAGATATATGGAAACCCCTGGATGCCCAGGCAGAAGTTTGCTGTAGGGGTGGGGT 38093

Qy     929 snwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccnc 988
           : :: |||:| || :: || :| || || ||:||:|| ||:|| || |||||:|| || |
Db   38094 CCTCATGGAGAACCTCTGCAAGGGTAGTACAAAAGGGAAATGTTGGGTGGGAGCCCCCAC 38153

Qy     989 aymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnm 1048
           |:|| || || :: || || || :: :: :| || || :| :| :: || || :: :: :
Db   38154 ACAGAGTCCCCAGTGGGGCTCCATCTAGTAGAGCTGTGAGAAGAAGTCCACCATCCTCCA 38213

Qy    1049 gnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngaya 1108
           | :| ||:||:|| :| :: || ||::: :| ||:||:|| || ||:||::: || ||:|
Db   38214 GACTCCAGAAGGGTAGATCCACTGACAGCTTGCAGCATGTGCCTGAAAAATCCACAGACA 38273

Qy    1109 cncartgycarccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtng 1168
           | ||:||:||:|| || ||:|| || || |||||::: || || ||:||:|| || || |
Db   38274 CTCAGTGCCAGCCTGTGAAAGCAGCAGGGATGGAGTCTGTACCCTACAAAACCGTAGTGG 38333

Qy    1169 cngarytnacnaaracngtnggnathtayytnytncaytgycaygayytngaygtnmgnc 1228
           | ||::| || ||:|| || || ||:||::| :| ||:||:||:||::| ||:|| :| |
Db   38334 CAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGCATTGTCATGACCTGGACGTGAGAC 38393

Qy    1229 ayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgna 1288
           |:|| || ||::| ||:||:||:|| || :| :| ||:||:||:|| || || ||::| |
Db   38394 ATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCCACTGGATTTCGGA 38453

Qy    1289 cntayatgggnccngtnccnytntgyttyggncarttyttyccnttyggnacngcngtnt 1348
           | ||:|||| || || || :| ||:||:|| ||:||:||:|| ||:|| || || || |
Db   38454 CTTATATGGGGCCCGTACCCCTTTGTTTTGGCCAATTTTTTCCATTTGGAACTGCCGTAT 38513

Qy    1349 tyacncartgyytntayytncaytgyatg 1377
           |:|| ||:||::| ||::| ||:||:|||
Db   38514 TTACCCAATGCCTGTACCTCCATTGTATG 38542


RESULT   2
AC025577
LOCUS       AC025577              154090 bp    DNA     linear   PRI 25-AUG-2000
DEFINITION  Homo sapiens 12 BAC RP11-13C3 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC025577
VERSION     AC025577.15  GI:9910028
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154090)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
```

Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Hohnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Muzny,D., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Nelson,D. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 154090)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 154090)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Aug 25, 2000 this sequence version replaced gi:9664948.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

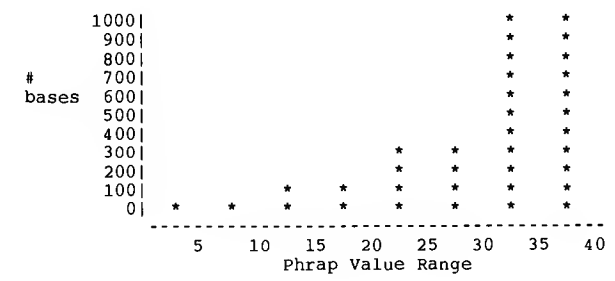
---------- Summary Statistics ----------

Contig length:	154090
Phrap values in estimate:	153545
Average error rate (BCM-Phrap estimate):	0.000261755
Fraction of Phrap values less than 40 :	0.0284086
Number of consensus changing edits:	22
Number of N's in consensus :	0

---------- Consensus changing edits ----------

Position	Original+Context	Edited+Context
9347	aacacagacc(n)ttttttttt	aacacagacc(t)ttttttttt
10135	atttgcccct(n)taggaaaaga	atttgcccct(t)taggaaaaga
10353	ctgtcaccag(n)aaaagatcca	ctgtcaccag(t)aaaagatcca
10417	agcaatgcag(n)ggctacagaa	agcaatgcag(t)ggctacagaa
61959	ctttactatt(n)tgtttattct	ctttactatt(t)gttttattct
61960	tttactattn(t)gtttattcta	tttactattt(g)ttttattcta
61961	ttactattnt(g)tttattctaa	ttactatttg(t)tttattctaa
61982	acaacacaag(n)ntaggtttgg	acaacacaag(a)ttaggtttgg
61983	caacacaagn(n)taggtttggn	caacacaaga(t)taggtttggt
61993	ntaggtttgg(n)ttactcttcc	ttaggtttgg(t)ttactcttcc
62205	tcattgaccc(n)ctggtcattc	tcattgaccc(a)ctggtcattc
62956	actgcaacct(n)tgcctcccag	actgcaacct(c)tgcctcccag
63722	tttacatata(n)cagtacagta	tttacatata(a)cagtacagta
90212	cctagaaaaa(n)gacnttttct	cctagaaaaa(t)gacttttttct
90216	gaaaaangac(n)ttttcttttt	gaaaaatgac(t)ttttcttttt
90228	tttcttttta(n)ttttttcnat	tttcttttta(t)tttttctat
90236	tantttttc(n)atctttaggg	tatttttttc(t)atctttaggg
90443	accttagatg(n)ctcctcccag	accttagatg(a)ctcctcccag
91405	cacagcttaa(n)caaggtaaga	cacagcttaa(a)caaggtaaga
122290	aaaaaaagaa(n)gtaagagaaa	aaaaaaagaa(a)gtaagagaaa
137090	acagagaaag(n)caaaacactc	acagagaaag(a)caaaacactc
141632	tgaagccact(n)gaaaagtaat	tgaagccact(t)gaaaagtaat

--------- Distribution of Quality < 40 Bases ----------



Version: 1.01 qxfo.

FEATURES Location/Qualifiers

```
     source          1. .154090
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-13C3"
     repeat_region   54. .288
                     /rpt_family="MIR"
     repeat_region   1020. .1146
                     /rpt_family="MER3"
     repeat_region   complement(1248. .1584)
                     /rpt_family="MER46B"
     repeat_region   complement(1624. .1682)
                     /rpt_family="MIR"
     repeat_region   2630. .2761
                     /rpt_family="GA-rich"
     repeat_region   4088. .4216
                     /rpt_family="L2"
     repeat_region   5232. .5301
                     /rpt_family="MER102"
     repeat_region   complement(5450. .5749)
                     /rpt_family="AluY"
     repeat_region   6571. .6616
                     /rpt_family="(TG)n"
     repeat_region   complement(6678. .7320)
                     /rpt_family="L1ME3"
     repeat_region   7419. .7454
                     /rpt_family="AT_rich"
     repeat_region   complement(7741. .7807)
                     /rpt_family="MER5A"
     repeat_region   8113. .8267
                     /rpt_family="MER5A"
     repeat_region   8385. .8587
                     /rpt_family="MIR"
     repeat_region   8589. .8630
                     /rpt_family="L2"
     repeat_region   8666. .8695
                     /rpt_family="(TTA)n"

  Query Match             36.2%;  Score 497.8;  DB 9;  Length 154090;
  Best Local Similarity   47.7%;  Pred. No. 9.6e-109;
  Matches  488;  Conservative  216;  Mismatches  282;  Indels   36;  Gaps    2;

Qy      390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
            :||:||:||::: ||: |:||::: ||: |:||:|| || | | : :| ||:|  ||
Db 101633 GGAAGAAATTTCTAAGTAGCAAAGCATTAAAGAGGTGACTTAGATGTGGTTAAAGGTATT 101692

Qy      450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
            ::: | :||::| ||:|| ||:||:||: : :  ||:||::| || || ||:||:|| ||
Db 101693 CAGTTATATAAGGGAGGCAGAGCATAAACGTTTGGAAAATTTGCAGCCTGACAATGTGAT 101752

Qy      510 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 569
            : |:||:||:||:|| ||::: ||:|| || || ||::| || || ||:||:||:||: |
Db 101753 AGAAAAGAAAAACCCATTTTCTGAGGGGAATTTGAAGCTGGCTGCAGAAATTTGCATAAG 101812

Qy      570 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 629
            :||: |:||::  ||:|| ||:|| ||:||:||:|| ||:||: |:::   | | | :||
Db 101813 TAACAAGGAGCCTAATGTTAATCCCCAAGACAATGGGGAAAATGTCTCCAAGGCATATCA 101872

Qy      630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 689
            ::| :: ::   :||: : ||:  ::: :  ||  ||:| || :| :| ||:|||||:||
Db 101873 GAGGTCTTCATGGCAGCCCATCCCATCACAGGCCCGGAGGCCTAGGAGAAAATGGTTTTG 101932

Qy      690 yggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt--- 746
            : |  | || || ||  : :  ||:||:|| ||:|| :| ||::| || || ||:||
Db 101933 TAGGCCAGGCCCAGGAGCCCCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGTGTATA 101992

Qy      747 ------------------------------nccngtnaaywsngcngtngcnwsngar 774
                                           |       ::: || || || :: ||:
Db 101993 CCAGCTGCTGCAGTGGTGGCTGAAAGGGGCCAGCATAGAGCTTGGGCCGTGGCTTCGGAG 102052

Qy      775 ggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraar 834
            || || :: || ||:|| |||||::| || :: |  || ||:|| |  || ||  |:||:
Db 102053 GGTGCAAGCCCCAAGCCTTGGCAATTTCCATGTGTTGTTGAGCCTGAGGGTGCACAGAAG 102112

Qy      835 wsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaaycentgg 894
            :: :| ||:||:|  |||||:|| ||   ::| ||:||:| :|| ||:|| ||: | |||
Db 102113 TCAAGAATTGAGGCTTGGGAGCCTCCACCTAGATTTCAGAGGATGTATGGAAATGCCTGG 102172

Qy      895 atgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtn 954
            ||| | :| ||:||:||:||    || || ||  : :: |||:| || :: || :| |
Db 102173 ATGTCCAGGCAGAAGTTTGCTTCAGGGGTGGGGCCCTCATGGAGAACCTCTGCTAGGGCA 102232

Qy      955 gtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsn 1014
            || ||:||:|| ||:|| || | |||:||    ||::| || || :: || || || ::
Db 102233 GTGCAGAAGGGAAATGTGGGGTTGGAGCCCATACACAGAGTCCCTACTGGGGCACCATCT 102292

Qy     1015 wsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngay 1074
            ::  | || || :| :|  : || || :: :: :| :  ||:|| || :| :: ||  |:
Db 102293 AGGGGAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAATGGTAGATCCACCAAC 102352

Qy     1075 wsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcn 1134
            :: :| || ||:|| || | :||: : || || || ||:  :||:||  | ||:|| ||
Db 102353 AGCTTGCACCATGTACCTGGAAAAGCCACAGAGACTCAACACCAGCCCATGAAAGCAGCC 102412

Qy     1135 ggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1194
              | |||: : || || | :||: |    |  || ||:|  | ||:||  | || | :
Db 102413 AAGAGGGAGCCTGTACCCTGCAAAGCCACAGAAGCAGAGCTTCCCAAGACCATGGGAACC 102472

Qy     1195 tayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggn 1254
             |::| :| ||:     :||::| ||:|| :  ||:|  || ||: | ||:||: |: |
Db 102473 CACCTCTTGCATCAGTGTGACCTAGATGTGAAACATGCAGTCAAAGGAGATCATATTAGA 102532

Qy     1255 gcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntg 1313
             | :| :| ||:||:| :||  | || ||::  || | :||||| |   |  || :| ||
Db 102533 ACTTTGAGATTTGACTACCCTTCTGGGTTTCAGACTTGCATGGGGCATATAGCCCCTTTG 102592

Qy     1314 yttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1373
            :||:|| ||:||: |:|| ||: | || || || ||:|| ||:||::| ||::  ||:||
Db 102593 TTTTGGCCAATTTCTCCCATTTAGAACAGCTGTATTTACTCAATGCCTGTATCGCCATTG 102652

Qy     1374 ya 1375
            :|
Db 102653 CA 102654


RESULT   3
AC026107
LOCUS       AC026107      .          186660 bp    DNA     linear   HTG 06-FEB-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-307L1, WORKING DRAFT
            SEQUENCE, 1 unordered pieces.
ACCESSION   AC026107
VERSION     AC026107.22  GI:13899178
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186660)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
```

```
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,1.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186660)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 1, 2001 this sequence version replaced gi:13877175.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HAMK
                Center clone name: RP11-307L1
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 18% of reads
                Chemistry: Dye-terminator Big Dye: 82% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 186630 bases at least Q40
                Consensus quality: 186659 bases at least Q30
                Consensus quality: 186660 bases at least Q20
                Estimated insert size: 187779; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 10.7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   186660: contig of 186660 bp in length.
FEATURES            Location/Qualifiers
     source         1. .186660
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="12"
                    /clone="RP11-307L1"
BASE COUNT    57089 a  37346 c  35962 g  56263 t
ORIGIN
```

```
  Query Match           36.2%;  Score 497.8;  DB 2;  Length 186660;
  Best Local Similarity  47.7%;  Pred. No. 1e-108;
  Matches  488;  Conservative  216;  Mismatches  282;  Indels   36;  Gaps    2;

Qy     390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
           :||:||:||::: ||: |:||::: ||: |:|||:|| || | | | :  :| ||:|  ||
Db   71882 GGAAGAAATTTCTAAGTAGCAAAGCATTAAAGAGGTGACTTAGATGTGGTTAAAGGTATT 71941

Qy     450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
           ::: | :||::|  |:|| ||:||:||: : :  ||:||::| || || ||:||:|| ||
Db   71942 CAGTTATATAAGGAAGGCAGAGCATAAACGTTTGGAAAATTTGCAGCCTGACAATGTGAT 72001

Qy     510 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 569
           : |:||:||:||:|| ||::: ||:|| || || ||::| || || ||:||:||:||: |
Db   72002 AGAAAAGAAAAACCCATTTTCTGAGGGGAATTTGAAGCTGGCTGCAGAAATTTGCATAAG 72061

Qy     570 yaaygargarytnaaygtnaaycccncargayaayggngaraayathwsntggacntgyca 629
           :||: |:||::  ||:|| ||:|| ||:||:||:|| ||:||: |:::   | | | :||
Db   72062 TAACAAGGAGCCTAATGTTAATCCCCAAGACAATGGGGAAAATGTCTCCAAGGCATATCA 72121

Qy     630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 689
           ::| :: ::   :||: : ||:  ::: :  ||  ||:| || :| :| ||:|||||:||
Db   72122 GAGGTCTTCATGGCAGCCCATCCCATCACAGGCCCGGAGGCCTAGGAGAAAATGGTTTTG 72181

Qy     690 yggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt--- 746
           : |  | || || ||  : :  ||:||:|| ||:|| :| ||::| || || ||:||
Db   72182 TAGGCCAGGCCCAGGAGCCCCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGTGTATA 72241

Qy     747 -------------------------------nccngtnaaywsngcngtngcnwsngar 774
                                           |      ::: || || || :: ||:
Db   72242 CCAGCTGCTGCAGTGGTGGCTGAAAGGGGCCAGCATAGAGCTTGGGCCGTGGCTTCGGAG 72301

Qy     775 ggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraar 834
           || || :: || ||:|| |||||::| || :: || || ||:|| |  || || |:||:
Db   72302 GGTGCAAGCCCCAAGCCTTGGCAATTTCCATGTGGTGTTGAGCCTGAGGGTGCACAGAAG 72361

Qy     835 wsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaaycccntgg 894
           :: :| ||:||:|  |||||:|| ||   ::| ||:||:| :|| ||:|| ||: | |||
Db   72362 TCAAGAATTGAGGCTTGGGAGCCTCCACCTAGATTTCAGAGGATGTATGGAAATGCCTGG 72421

Qy     895 atgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtn 954
           ||| | :| ||:||:||:||    || || ||  : :: |||:| || :: || :| |
Db   72422 ATGTCCAGGCAGAAGTTTGCTTCAGGGGTGGGGCCCTCATGGAGAACCTCTGCTAGGGCA 72481

Qy     955 gtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsn 1014
           || ||:||:|| ||:|| || | |||:||    ||::| || || :: || || || ::
Db   72482 GTGCAGAAGGGAAATGTGGGGTTGGAGCCCATACACAGAGTCCCTACTGGGGCACCATCT 72541

Qy    1015 wsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngay 1074
           ::  | || || :| :|  : || || :: :: :| :  ||:|| || :| :: ||  |:
Db   72542 AGGGGAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAATGGTAGATCCACCAAC 72601

Qy    1075 wsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcn 1134
           :: :| || ||:|| || | :||: : || || || ||:  :||:||  | ||:|| ||
Db   72602 AGCTTGCACCATGTACCTGGAAAAGCCACAGAGACTCAACACCAGCCCATGAAAGCAGCC 72661

Qy    1135 ggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1194
              | |||: : || || | :||: |    |  || ||::|  | ||:|| | || | :
Db   72662 AAGAGGGAGCCTGTACCCTGCAAAGCCACAGAAGCAGAGCTTCCCAAGACCATGGGAACC 72721

Qy    1195 tayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggn 1254
            |::| :| ||:     :||::| ||:|| :  ||:|  || ||: | ||:||: |: |
Db   72722 CACCTCTTGCATCAGTGTGACCTAGATGTGAAACATGCAGTCAAAGGAGATCATATTAGA 72781

Qy    1255 gcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntg 1313
            | :| :| ||:||:| :||  | || ||::  || | :|||||  |   |  || :| ||
Db   72782 ACTTTGAGATTTGACTACCCTTCTGGGTTTCAGACTTGCATGGGGCATATAGCCCCTTTG 72841

Qy    1314 yttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1373
           :||:|| ||:||: |:|| ||: | || || || ||:|| ||:||::| ||::  ||:||
Db   72842 TTTTGGCCAATTTCTCCCATTTAGAACAGCTGTATTTACTCAATGCCTGTATCGCCATTG 72901

Qy    1374 ya 1375
```

```
        :|
Db  72902 CA 72903


RESULT   4
AL138962/c
LOCUS       AL138962              124518 bp    DNA     linear   PRI 28-JAN-2001
DEFINITION  Human DNA sequence from clone RP11-279N8 on chromosome 13, complete
            sequence.
ACCESSION   AL138962
VERSION     AL138962.16  GI:12597036
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124518)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 29, 2001 this sequence version replaced gi:12584354.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; WP:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-279N8 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-279N8 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP11-279N8 is at 124518 in this
            sequence. The true left end of clone RP11-528D24 is at 114969 in
            this sequence. The true right end of clone RP11-214O11 is at 100 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1. .124518
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-279N8"
                     /clone_lib="RPCI-11.1"
     misc_feature    112388. .112496
                     /note="Big dye primer and dGTP big dye chemistries only."
BASE COUNT    39295 a  26385 c  24794 g  34044 t
ORIGIN


  Query Match           35.8%;  Score 493.6;  DB 9;  Length 124518;
  Best Local Similarity   47.5%;  Pred. No. 9.2e-108;
  Matches  486;  Conservative  216;  Mismatches  285;  Indels   36;  Gaps    2;

Qy    390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
          :||:|| ||::: ||:||: |::: ||:||:||:|| || ||||| :| :|  |:|| |

Db  66716 GGAAGACATTTCTAAGCAGAAAAGCATTCAAGAGGTGACTTGGGTACTGTTAGAGGCATC 66657

Qy    450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
          ::: ||:|||::| ||: | ||: |:||::: :   |:|||::| || || ||:||:|| ||
Db  66656 CAGTTTTATAAGGGAAACAGAGTATAAAAGTTTGAAAAATTTGCAGCCTGACAATGTGAT 66597

Qy    510 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 569
          : |:|||:||:| :|| ||::: ||:|  ||: |:|||::| || || ||:||:||:||: |
Db  66596 AGAAAAGAAAATCCCATTTTCTGAGGAGAAAGTCAAGCTGGCTGCAGAAATTTGCATAAG 66537

Qy    570 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 629
          :||:||:||::  ||:|| ||:|  ||:||:||:|| | :||: |:::  || | ||:||
Db  66536 TAACGAGGAGCCAAATGTTAATCACCAAGACAATGGAGGAAATGTCTCCAGGGCATGTCA 66477

Qy    630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naartggtt 686
          ::| :: :: ||:||  :  |:  ::: :  || |||:  |  :| :|    ||:|||||
Db  66476 GAGGTCATCACAGCATCCCCTCCCATCACAGGCCTGGAAGCTTAGGAGGAAAAAATGGTT 66417

Qy    687 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 746
          :  :|| |  ||  | || :: :| ||:||:|| ||:|| :| ||:::| ||  | ||: |
Db  66416 TCATGGAAGGGGCACAGGGTCCCTGTGTTGTGTGCAGCCTAGGGACTTGGTGTCCTGCAT 66357

Qy    747 nccngtnaayws----------------------------------ngcngtngcnwsnga 773
            || |  : :                                    || || || :: ||
Db  66356 CCCAGCTGCTCCAGCTGTGGCTGAAAGGGGCCAATGTGGAGTTCAGGCTGTGGCTTCAGA 66297

Qy    774 rggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraa 833
          :|| || :: || ||:|| |||||::| ||    |  ||  |:|| ||  | ||  |:||
Db  66296 GGGTGCAAGCCCTAAGCCTTGGCAGCTTCCACATGATGTTAAGCCTGTGAGTGCACAGAA 66237

Qy    834 rwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaaycentg 893
          ::: :| | :| :|| |||||:|| |    ::| ||:||:| :||:||:|| ||: | ||
Db  66236 ATCAAGAACTGGGGTTTGGGAACCTCTGTCTAGATTTCAGAGGATATATGGTAATACCTG 66177

Qy    894 gatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngt 953
          |||||| :| ||:||:||:|| ||  | |  ||    :: || :  || :: || :| |
Db  66176 GATGCCCAGACAGAAGTTTGCTGTATGGGAGGGGGTCTCATGAAAAACCTCTGCTAGGGC 66117

Qy    954 ngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnws 1013
            ||   : |:|| ||:|| || |  ||:|   | ||::| || || :: || || ||  :
Db  66116 AGTGTGGGAGGGAAATGTGGGGTCAGAGCTGTCACACAGAGTCCCTACTGGGGCACCGCC 66057

Qy   1014 nwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacnga 1073
            ::  | || || :  :|  : || |  :: :: :| :  ||:|| || :| :: || ||
Db  66056 TAGTGGAGCTGTGAAAAGTGGGCCACTATCCTCCAGACCCCAGAATGGTAGATCCACTGA 65997

Qy   1074 ywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngc 1133
          ::: :| || ||:|  || | :||: : || ||:|| ||:||:||:|| || ||:|| ||
Db  65996 CAGCTTGCACCATGCACCTGGAAAAGCCACAGACACTCAATGCCAGCCTGTGAAAGCAGC 65937

Qy   1134 nggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnat 1193
            || | |||::: || || | :||:||    |  || ||::|  | ||:||  | || |
Db  65936 TGGGAGGGAGACTGTACCATGCAAGACCACAGGGGCAGAGCTGCCCAAGACCATGGGAAC 65877

Oy   1194 htayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttygg 1253
          : |::| :| ||:   ||:|||::| ||:|| :| ||:|| || ||: | ||:||:||:||
Db  65876 CCACCTCTTGCATCAGCATGACCTGGATGTGAGACATGGAGTCAAAGGAGATCATTTTGG 65817

Qy   1254 ngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytntg 1313
            || :| :| ||:||:| :||  | || ||::  || | :|||||  |     || :| ||
Db  65816 AGCTTTAAGATTTGACTTCCCTGCTGGATTTCAGACTTGCATGGGCCTGTAGCCCCTTTG 65757

Qy   1314 yttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1373
          :||:|| ||:||:  :|| ||:|| |  |   | ||:|| ||:||::| ||::  ||:||
Db  65756 TTTTGGCCAATTTCCCCCATTTGGAATGGGTATATTTACCCAATGCCTGTACCCCCATTG 65697

Qy   1374 yat 1376
          :||
Db  65696 TAT 65694


RESULT   5
AC008799
```

```
LOCUS       AC008799              123070 bp    DNA     linear   HTG 18-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone CTD-2061E19, WORKING DRAFT
            SEQUENCE, 8 ordered pieces.
ACCESSION   AC008799
VERSION     AC008799.4  GI:9256046
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 123070)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 123070)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:7709316.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 651088
            Center clone name: CITB-H1_2061E19
            -------
            Summary Statistics
            Consensus quality: 116295 bases at least Q40
            Consensus quality: 121288 bases at least Q30
            Consensus quality: 122086 bases at least Q20
            Estimated insert size: 123000; pulse field gel estimation
            Estimated insert size: 122770; sum-of-contigs estimation
            Quality coverage:  6.04 in Q20 bases; pulse field gel estimation
            Quality coverage:  6.05 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     9436: contig of 9436 bp in length
            *     9437     9536: gap of unknown length
            *     9537    26537: contig of 17001 bp in length
            *    26538    26637: gap of unknown length
            *    26638    45958: contig of 19321 bp in length
            *    45959    46058: gap of unknown length
            *    46059    68657: contig of 22599 bp in length
            *    68658    68757: gap of unknown length
            *    68758    72152: contig of 3395 bp in length
            *    72153    72252: gap of unknown length
            *    72253    81069: contig of 8817 bp in length
            *    81070    81169: gap of unknown length
            *    81170   121547: contig of 40378 bp in length
            *   121548   121647: gap of unknown length
            *   121648   123070: contig of 1423 bp in length.
FEATURES             Location/Qualifiers
     source          1. .123070
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2061E19"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    37763 a  24223 c  23855 g  36528 t    701 others
ORIGIN
```

Query Match 35.8%; Score 493; DB 2; Length 123070;

Best Local Similarity 46.7%; Pred. No. 1.3e-107;
Matches 495; Conservative 220; Mismatches 309; Indels 36; Gaps 2;

```
Qy     352 aayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcar 411
           ||: |:|  :|    :| ||  |  |:           :||:||:||::: ||:||:||:
Db  107907 AACTTTTAACTTGAAAGAAATGATTTAGGGTATCTGGCAGAAGAAATTTCTAAGCAGCAA 107966

Qy     412 wsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngar 471
           :: ||:| :||:|| || ||||| :| :| ||:|  ||::: ||:||::| ||:|| ||:
Db  107967 AGCATTCCAGAGGTGACTTGGGTGCTGTTAAAGGTATTCAGTTTTATAAGGGAAGCAGAG 108026

Qy     472 cayaarwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccnttywsn 531
           ||:||:::    ||:||::| ||  | ||:||:|| ||: |:||:||:| :|| ||:::
Db  108027 CATAAAAGTCTGGAAAATTTGCAGTCCGACAATGTGATAGAAAAGAAAATTCCATTTTCT 108086

Qy     532 garggnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaay 591
           ||:|  ||:||:||::| || || ||:|| ||:||: |:||: |:||::| ||: | ||:
Db  108087 GAGGAGAAATTCAAGCTGGCTGCAGAAATGTGCATAAGTAACAAGGAGCTGAATATTAAT 108146

Qy     592 ccncargayaaygggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnath 651
           || ||:||:||:|| ||:||:||:::   | | ||:||::| :: :: | :||: :  |:
Db  108147 CCCCAAGACAATGGGGAAAACATCTCCAAGGCATGTCAGAGGTCTTCACGGCAGCCCCTC 108206

Qy     652 aarwsnytngcntggmgnccnmgnmg---naartggttytgyggnacnggnccnggnwsn 708
             ::: :   |    :| || :| :|    ||:|||||:||:|| | || || || ::
Db  108207 CCATCATAAACCCAAAGGCCTAGGAGGAAAAAGTGGTTTTGTGGGCCAGGCCCAGGGTCC 108266

Qy     709 ytntgytgygtncarccnmgngayytngtnccntgygtnccngtnaa---------- 755
           :| ||:||:|| ||:|  :| ||::| || || ||: | || |
Db  108267 TTGTGCTGTGTGCAGCTTAGGGACTTGGTGCCCTGCATCCCAGCTGCTCCAGCCATGGCT 108326

Qy     756 --------------------ywsngcngtngcnwsngarggngcnwsnccnaarccntgg 795
                               : : || || || :: ||:|| || :: || ||:|| |||
Db  108327 GAAAGGGGCCAACATAGAGCTCGGGCTGTGGCTTCAGAAGGTGCAAGCCCCAAGCCTTGG 108386

Qy     796 carytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggar 855
           ||:|| || :: || || ||:|| |   | ||  |:||::: :  ||:| :|| |||||:
Db  108387 CAGCTTCCATGTGGTGTTGAGCCTGCAAGTGCACAGAAGTCACAAATTGGGGTTTGGGAA 108446

Qy     856 ccnccnathmgnttycaraarathtayggnaayccntggatgccnmgncaraarttygcn 915
           || ||   ::| || ||:| :|| ||:|| ||: |  |||| | :| ||:||:||:||
Db  108447 CCTCCACCTAGATTGCAGATGATGTATGGAAATGCCCGGATGACCAGGCAGAAGTTTGCT 108506

Qy     916 gtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggn 975
           |  || |  ||  : :: |||:| || :: || :| |  || ||:||:   ||:|| ||
Db  108507 GCAGGGGCAGGGCCTTCATGGAGAACCTCTGCTAGGGCAGTGCAGAAGAAAAATGTAGGG 108566

Qy     976 tgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsn 1035
           |  ||:||  | ||:| || || ::  | || |   : ::  | || || :| :| ::
Db  108567 TCAGAGCCCACACACAGAGTACCCACTAGAGCACTGCCTAGTGGAGCTGTGAGAAGAAGG 108626

Qy    1036 ccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngar 1095
           || || :: :: :| :| ||:|| || :| :: ||  |::: :| ||   :|| || | :
Db  108627 CCACCATCCTCCAGACTCCAGAATGGTAGGTCCACCAACAGCTTGCACTGTGTACCTGGA 108686

Qy    1096 aarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntay 1155
           ||: : || | :|| || ||:||:|| || ||:|| ||  | | |||: :  | || | :
Db  108687 AAAGCCACAGGCACTCATTGCCAGCCTGTGAAAGCAGCCAGGAGGGAGGCTATCCCCTGC 108746

Qy    1156 aaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygay 1215
           ||:      |  ||  |::|  | ||:||  | || | :||::| :| | :   | :||:
Db  108747 AAAGTCACAGGGGCGAAGCTGTCCAAGACCTTGGGAACCTACCTCTTGCCTCAGCGTGAC 108806

Qy    1216 ytngaygtnmgncaygggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccn 1275
           :  ||:|| :| ||:|| || ||: | ||:||:||:|| || :| :| ||:||:||:||
Db  108807 CCGGATGTGAGACATGGAGTCAAAGGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCT 108866

Qy    1276 acnggnttymgnacntayatgggnccngtnccnytntgyttyggncarttyttyccntty 1335
            | || ||::  ||      ||| |     || :| ||:||:|| ||:||: |:|| ||:
Db  108867 GCTGGATTTCAGACTGCATGGGGCCTGTAGCCCCTTTGCTTTGGCCAATTTCTCCCATTT 108926

Qy    1336 ggnacngcngtnttyacncartgyytntayytncaytgya 1375
           || || || || ||:|| | :||::| ||::   |:||:|
```

```
Db 108927 GGAACAGCTGTATTTACCCCATGCCTGTACCCCGATTGTA 108966


RESULT   6
AC044889/c
LOCUS       AC044889              189768 bp    DNA     linear   HTG 22-MAY-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-792C1 map 2, WORKING DRAFT
            SEQUENCE, 36 unordered pieces.
ACCESSION   AC044889
VERSION     AC044889.2  GI:8016676
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189768)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 2, clone RP11-792C1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189768)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 22, 2000 this sequence version replaced gi:7543856.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9594
                Center clone name: 792_C_1
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 168749 bases at least Q40
                Consensus quality: 180047 bases at least Q30
                Consensus quality: 184014 bases at least Q20
                Insert size: 190000; agarose-fp
                Insert size: 186268; sum-of-contigs
                Quality coverage: 3.8 in Q20 bases; agarose-fp
                Quality coverage: 3.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 36 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1015: contig of 1015 bp in length
            *    1016 1115: gap of      100 bp
            *    1116     2284: contig of 1169 bp in length
            *    2285 2384: gap of      100 bp
            *    2385     3898: contig of 1514 bp in length
            *    3899 3998: gap of      100 bp
            *    3999     5259: contig of 1261 bp in length
            *    5260 5359: gap of      100 bp
            *    5360     6912: contig of 1553 bp in length
            *    6913 7012: gap of      100 bp
            *    7013     8145: contig of 1133 bp in length
            *    8146 8245: gap of      100 bp
            *    8246     9694: contig of 1449 bp in length
            *    9695 9794: gap of      100 bp
            *    9795    11004: contig of 1210 bp in length
            *   11005 11104: gap of      100 bp
            *   11105    12907: contig of 1803 bp in length
            *   12908 13007: gap of      100 bp
            *   13008    14916: contig of 1909 bp in length
            *   14917 15016: gap of      100 bp
            *   15017    17114: contig of 2098 bp in length
            *   17115 17214: gap of      100 bp
            *   17215    18527: contig of 1313 bp in length
            *   18528 18627: gap of      100 bp
            *   18628    20561: contig of 1934 bp in length
            *   20562 20661: gap of      100 bp
            *   20662    22484: contig of 1823 bp in length
            *   22485 22584: gap of      100 bp
            *   22585    25287: contig of 2703 bp in length
            *   25288 25387: gap of      100 bp
            *   25388    27603: contig of 2216 bp in length
            *   27604 27703: gap of      100 bp
            *   27704    30020: contig of 2317 bp in length
            *   30021 30120: gap of      100 bp
            *   30121    33472: contig of 3352 bp in length
            *   33473 33572: gap of      100 bp
            *   33573    38084: contig of 4512 bp in length
            *   38085 38184: gap of      100 bp
            *   38185    41069: contig of 2885 bp in length
            *   41070 41169: gap of      100 bp
            *   41170    45266: contig of 4097 bp in length
            *   45267 45366: gap of      100 bp
            *   45367    51092: contig of 5726 bp in length
            *   51093 51192: gap of      100 bp
            *   51193    55262: contig of 4070 bp in length
            *   55263 55362: gap of      100 bp
            *   55363    60982: contig of 5620 bp in length
            *   60983 61082: gap of      100 bp
            *   61083    67628: contig of 6546 bp in length
            *   67629 67728: gap of      100 bp
            *   67729    72690: contig of 4962 bp in length
            *   72691 72790: gap of      100 bp
            *   72791    79286: contig of 6496 bp in length
            *   79287 79386: gap of      100 bp
            *   79387    83451: contig of 4065 bp in length
            *   83452 83551: gap of      100 bp
            *   83552    92170: contig of 8619 bp in length
            *   92171 92270: gap of      100 bp
            *   92271    96333: contig of 4063 bp in length
            *   96334 96433: gap of      100 bp
            *   96434   103218: contig of 6785 bp in length
            *  103219 103318: gap of      100 bp
            *  103319   112553: contig of 9235 bp in length
            *  112554 112653: gap of      100 bp
            *  112654   123239: contig of 10586 bp in length
            *  123240 123339: gap of      100 bp
            *  123340   137921: contig of 14582 bp in length
            *  137922 138021: gap of      100 bp
            *  138022   157913: contig of 19892 bp in length
            *  157914 158013: gap of      100 bp
```

```
          *   158014   189768: contig of 31755 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189768
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-792C1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1015
                     /note="assembly_fragment"
     misc_feature    1116. .2284
                     /note="assembly_fragment"
     misc_feature    2385. .3898
                     /note="assembly_fragment"
     misc_feature    3999. .5259
                     /note="assembly_fragment"
     misc_feature    5360. .6912
                     /note="assembly_fragment"
     misc_feature    7013. .8145
                     /note="assembly_fragment"
     misc_feature    8246. .9694
                     /note="assembly_fragment"
     misc_feature    9795. .11004
                     /note="assembly_fragment"
     misc_feature    11105. .12907
                     /note="assembly_fragment"
     misc_feature    13008. .14916
                     /note="assembly_fragment"
     misc_feature    15017. .17114
                     /note="assembly_fragment"
     misc_feature    17215. .18527
                     /note="assembly_fragment"
     misc_feature    18628. .20561
                     /note="assembly_fragment"
     misc_feature    20662. .22484
                     /note="assembly_fragment"
     misc_feature    22585. .25287
                     /note="assembly_fragment"
     misc_feature    25388. .27603
                     /note="assembly_fragment"
     misc_feature    27704. .30020
                     /note="assembly_fragment"
     misc_feature    30121. .33472
                     /note="assembly_fragment"
     misc_feature    33573. .38084
                     /note="assembly_fragment"
     misc_feature    38185. .41069
                     /note="assembly_fragment"
     misc_feature    41170. .45266
                     /note="assembly_fragment"
     misc_feature    45367. .51092
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    51193. .55262
                     /note="assembly_fragment"
     misc_feature    55363. .60982
                     /note="assembly_fragment"

  Query Match             35.8%;  Score 493;  DB 2;  Length 189768;
  Best Local Similarity   46.7%;  Pred. No. 1.5e-107;
  Matches  495;  Conservative  220;  Mismatches  309;  Indels   36;  Gaps    2;

Qy     352 aayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcar 411
           ||: |:|  :|    :| ||  |  |:          :||:||:||::: ||:||:||:
Db   86595 AACTTTTAACTTGAAAGAAATGATTTAGGGTATCTGGCAGAAGAAATTTCTAAGCAGCAA 86536

Qy     412 wsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngar 471
           :: ||:| :||:|| || ||||| :| :| ||:|  ||::: ||:||::| ||:|| ||:
Db   86535 AGCATTCCAGAGGTGACTTGGGTGCTGTTAAAGGTATTCAGTTTTATAAGGGAAGCAGAG 86476

Qy     472 cayaarwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccnttywsn 531
           ||:||:::    ||:||::| ||  | ||:||:|| ||: |:||:||:| :|| ||:::
Db   86475 CATAAAAGTCTGGAAAATTTGCAGTCCGACAATGTGATAGAAAAGAAAATTCCATTTTCT 86416

Qy     532 garggnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaay 591
           ||:|  ||:||:||::| || || ||:|| ||:||: |:||: |:||::| ||: | ||:
Db   86415 GAGGAGAAATTCAAGCTGGCTGCAGAAATGTGCATAAGTAACAAGGAGCTGAATATTAAT 86356

Qy     592 ccncargayaayggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnath 651
           || ||:||:||:|| ||:||:||:::   | | ||:||::| :: :: | :||: :  |:
Db   86355 CCCCAAGACAATGGGGAAAACATCTCCAAGGCATGTCAGAGGTCTTCACGGCAGCCCCTC 86296

Qy     652 aarwsnytngcntggmgnccnmgnmg---naartggttytgyggnacnggnccnggnwsn 708
             ::: :  |   :| || :| :|    ||:|||||:||:||  | || || || ::
Db   86295 CCATCATAAACCCAAAGGCCTAGGAGGAAAAAGTGGTTTTGTGGGCCAGGCCCAGGGTCC 86236

Qy     709 ytntgytgygtncarccnmgngayytngtnccntgygtnccngtnaa------------- 755
           :| ||:||:|| ||:|  :| ||::| || || ||: | || |
Db   86235 TTGTGCTGTGTGCAGCTTAGGGACTTGGTGCCCTGCATCCCAGCTGCTCCAGCCATGGCT 86176

Qy     756 --------------------ywsngcngtngcnwsngarggngcnwsnccnaarccntgg 795
                               : : || || || :: ||:|| || :: || ||:|| |||
Db   86175 GAAAGGGGCCAACATAGAGCTCGGGCTGTGGCTTCAGAAGGTGCAAGCCCCAAGCCTTGG 86116

Qy     796 carytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggar 855
           ||::| || :: || || ||:|| |   | ||  |:||::: :  ||:| :|| |||||:
Db   86115 CAGCTTCCATGTGGTGTTGAGCCTGCAAGTGCACAGAAGTCACAAATTGGGGTTTGGGAA 86056

Qy     856 ccnccnathmgnttycaraarathtayggnaaycntggatgccnmgncaraarttygcn 915
           || ||   ::| || ||:| :|| ||:|| ||: |  ||||| | :| ||:||:||:||
Db   86055 CCTCCACCTAGATTGCAGATGATGTATGGAAATGCCCGGATGACCAGGCAGAAGTTTGCT 85996

Qy     916 gtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggn 975
           |  || |  ||  : :: |||:| || :: || :| |  || ||:||:   ||:|| ||
Db   85995 GCAGGGGCAGGGCCTTCATGGAGAACCTCTGCTAGGGCAGTGCAGAAGAAAAATGTAGGG 85936

Qy     976 tgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsn 1035
           |  ||:||  | ||::| || || ::  | || |   : ::  | || || :| :| ::
Db   85935 TCAGAGCCCACACACAGAGTACCCACTAGAGCACTGCCTAGTGGAGCTGTGAGAAGAAGG 85876

Qy    1036 ccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngar 1095
           || || :: :: :| :| ||:|| || :| :: ||  |::: :| ||   :|| || | :
Db   85875 CCACCATCCTCCAGACTCCAGAATGGTAGGTCCACCAACAGCTTGCACTGTGTACCTGGA 85816

Qy    1096 aarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntay 1155
           ||: : || | :|| || ||:||:|| || ||:|| ||  | | |||: :  | || | :
Db   85815 AAAGCCACAGGCACTCATTGCCAGCCTGTGAAAGCAGCCAGGAGGGAGGCTATCCCCTGC 85756

Oy    1156 aaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygay 1215
           ||:      |  ||  |::|  | ||:||  | || | :||::| :| | :  | :||:
Db   85755 AAAGTCACAGGGGCGAAGCTGTCCAAGACCTTGGGAACCTACCTCTTGCCTCAGCGTGAC 85696

Qy    1216 ytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccn 1275
           :  ||:|| :| ||:|| || ||: | ||:||:||:|| || :| :| ||:||:||:||
Db   85695 CCGGATGTGAGACATGGAGTCAAAGGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCT 85636

Qy    1276 acnggnttymgnacntayatgggnccngtnccnytntgyttyggncarttyttyccntty 1335
           | || ||::  ||      ||| |     || :| ||:||:|| ||:||: |:|| ||:
Db   85635 GCTGGATTTCAGACTGCATGGGGCCTGTAGCCCCTTTGCTTTGGCCAATTTCTCCCATTT 85576

Qy    1336 ggnacngcngtnttyacncartgyytntayytncaytgya 1375
           || || || || ||:|| | :||::| ||::   |:||:|
Db   85575 GGAACAGCTGTATTTACCCCATGCCTGTACCCCGATTGTA 85536


RESULT   7
AC017063
LOCUS       AC017063              198545 bp    DNA     linear   HTG 02-NOV-2001
DEFINITION  Homo sapiens chromosome 4 clone RP11-354H17, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC017063
VERSION     AC017063.7  GI:16596637
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198545)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198545)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 2, 2001 this sequence version replaced gi:15741601.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_NH0354H17
            ------------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 35%
            Sequencing vector: plasmid; 65%
            Chemistry: Dye-primer ET; 34% of reads
            Chemistry: Dye-terminator Big Dye; 66% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 197437 bases at least Q40
            Consensus quality: 198145 bases at least Q30
            Consensus quality: 198648 bases at least Q20
            Insert size: 200000; agarose-fp
            Insert size: 200801; sum-of-contigs
            Quality coverage: 10.58 in Q20 bases; agarose-fp
            Quality coverage: 10.55 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1174: contig of 1174 bp in length
            *     1175     1274: gap of unknown length
            *     1275     2716: contig of 1442 bp in length
            *     2717     2816: gap of unknown length
            *     2817     4952: contig of 2136 bp in length
            *     4953     5052: gap of unknown length
            *     5053   198545: contig of 193493 bp in length.
FEATURES             Location/Qualifiers
     source          1. .198545
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-354H17"
     misc_feature    1. .1174
                     /note="assembly_name:Contig24"
     misc_feature    1275. .2716
                     /note="assembly_name:Contig38"
     misc_feature    2817. .4952
                     /note="assembly_name:Contig42"
     misc_feature    5053. .198545
                     /note="assembly_name:Contig44
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    62963 a  36710 c  37248 g  61324 t    300 others
ORIGIN

  Query Match             35.6%;  Score 490.4;  DB 2;  Length 198545;
  Best Local Similarity   47.0%;  Pred. No. 6.5e-107;
  Matches  498;  Conservative  213;  Mismatches  314;  Indels   34;  Gaps    2;

Qy    352 aayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcar 411
          ||: |:|  :|   |:|  |  |  |:           :||: |:||::: ||:||:||:
Db  83103 AACTTTTAACTTGGGAGAGATGATTTAGGGTATCTGGCAGAATAAATTTCTAAGCAGCAA 83162

Qy    412 wsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngar 471
          :: ||:||:|| || || ||||| :| :| ||:|| ||::: ||:||::| ||:|| ||:
Db  83163 AGCATTCAAGACGTGACTTGGGTGCTGTTAAAGGCATTCAGTTTTATAAGGGAAGCAGAG 83222

Qy    472 cayaarwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccnttywsn 531
          ||:||::  :  ||:||::| || || ||:||:|| ||: |:||:||:||:|| ||:::
Db  83223 CATAAAAATTTGGAAAATTTGCAGCCTGACAATGTGATAGAAAAGAAAAACCCATTTTCT 83282

Qy    532 garggnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaay 591
          ||:|  ||:|| ||::| || || ||:||:| :||: |:||:||:|| :| ||:|| ||:
Db  83283 GAGGAGAAATTGAAGCTGGCTGCAGAAATTTACATAAGTAATGAGGACCTGAATGTTAAT 83342

Qy    592 ccncargayaayggngaraayathwsntggacntgycarmgnwsnwscncarcarwsnath 651
          || ||:||:||:|| ||:||: |:::     | ||:||::| :: :  ||:||: :  |:
Db  83343 CCCCAAGACAATGGGGAAAATGTCTCCAAAGCATGTCAGAGGTCTTAACAGCAGCCCCTA 83402

Qy    652 aarwsnytngcntggmgnccnmgnmgnaartggttytgyggnacnggnccnggnwsnytn 711
            ::: :  || || :| || :| :| ||:||||||:| :||  | || || || :: :|
Db  83403 CCATCACAGGCCTGAAGGCCTAGGAGAAAATGGTTTTTCGGGCCAGGCCCAGGGTCCCTG 83462

Qy    712 tgytgygtncarccnmgngayytngtnccntgygtnccngtnaayws---------- 758
          ||:||:|| ||: | :|  |::| || || ||: | |  |  | : :
Db  83463 TGCTGTGTGCAGTCTAGGTACTTGGTGCCCTGCATCCTAGCCACTCCAGACATGACTAAA 83522

Qy    759 --------------------ngcngtngcnwsngarggngcnwsnccnaarccntggcar 798
                               || || || :: ||:|| |  ::    ||:|| |||| :
Db  83523 AGGGGCCAAGGTACAGCTCAGGCTGTTGCTTCAGAGGGTGGAAGCTGCAAGCCTTGGCGG 83582

Qy    799 ytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarccn 858
          :| || :  ||  | ||:|| |  ||  |  |:||::: :| ||: |:|| |||||:||
Db  83583 CTTCCAAATGGTATTGAACCTGCGGGTTCACAGAAGTCAAGAATTAAGGTTTGGGAACCT 83642

Qy    859 ccnathmgnttycaraarathtayggnaayccntggatgccnmgncaraarttygcngtn 918
          ||   ::| ||:||:||:|| ||:|| ||: | |||||||| :| ||:|| ||:|| |
Db  83643 CCGCCTAGATTTCAGAAAATGTATGGAAATGCCTGGATGCCCAGGCAGAATTTTGCTGCA 83702

Qy    919 ggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntgg 978
          || || ||  : :: |||:| || :: || :| |  || ||:||:|| ||:||  | |
Db  83703 GGAGTGGGGCCCTCATGGAGAACCTCTGCTAGGGCAGTGCAGAAGGGAAATGTGAGGTCA 83762

Qy    979 garccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccn 1038
          ||:|| || ||::| |  || :: || || ||  : ::  | || || :  :|  :  |
Db  83763 GAGCCCCCACACAGAGCCCCTACTGGGGCACCACCTAGTGGAGCTGTGAAAAGAGGGACA 83822

Qy   1039 ccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraar 1098
          |  :: :: :| :  ||:|| || :| :: || || :: :| || ||:|  |  | :||:
Db  83823 CTGTCCTCCAGACCCCAGAATGGTAGATCCACTGAGAGCTTGCACCATGCACTTGGAAAA 83882

Qy   1099 wsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntayaar 1158
            : || ||:||  |:  :||: |  | ||:|| ||  | | || : : || || ||:||:
Db  83883 GCCACAGACACTGAACACCAGACCATGAAAGCAGCCAGGAGGGTGGCTGTACCCTACAAA 83942

Qy   1159 acngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygayytn 1218
            |    |  |   |::|  | ||:||    || | : |::| :| ||:   ||:||::|
Db  83943 GCCACAGGGGTGAAGTTGGCCAAGACCACAGGAACCCACCTCTTGCATCAGCATGACTTG 84002

Qy   1219 gaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacn 1278
          ||:|| :| ||:|| || ||: |  |: |:||:|| || :| :| ||:||:||:|| ||
Db  84003 GATGTGAGACATGGTGTCAAAGGACATTATTTTGGAGCTTTAAGATTTGACTGCCCCACT 84062

Qy   1279 ggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnttygg 1337
          || ||: | || | :| ||| |  ||  || :| ||:|| || ||:| : |: | ||:||
Db  84063 GGATTTTGGACTTGCACGGGGCTTGTAACCCCTTTGTTTGGGCCAATGTCTCTCATTTGG 84122

Qy   1338 nacngcngtnttyacncartgyytntayytncaytgyat 1376
               || || ||:|  ||:||::| ||::  || ||: |
```

```
Db  84123 ATTTGCTGTATTTATCCAATGCCTGTACTGCCAATGTTT 84161


RESULT   8
AC025233/c
LOCUS       AC025233               161575 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-333E1, WORKING DRAFT
            SEQUENCE, 27 unordered pieces.
ACCESSION   AC025233
VERSION     AC025233.4  GI:8954317
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161575)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161575)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 7, 2000 this sequence version replaced gi:7801489.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0333E01
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 90%
            Sequencing vector: plasmid; 10%
            Chemistry: Dye-primer ET; 81% of reads
            Chemistry: Dye-terminator Big Dye; 19% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 144558 bases at least Q40
            Consensus quality: 149743 bases at least Q30
            Consensus quality: 152625 bases at least Q20
            Insert size: 176000; agarose-fp
            Insert size: 158975; sum-of-contigs
            Quality coverage: 3.48 in Q20 bases; agarose-fp
            Quality coverage: 3.75 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1141: contig of 1141 bp in length
            *     1142     1241: gap of unknown length
            *     1242     2424: contig of 1183 bp in length
            *     2425     2524: gap of unknown length
            *     2525     3961: contig of 1437 bp in length
            *     3962     4061: gap of unknown length
            *     4062     5712: contig of 1651 bp in length
            *     5713     5812: gap of unknown length
            *     5813     8019: contig of 2207 bp in length
            *     8020     8119: gap of unknown length
            *     8120    10208: contig of 2089 bp in length
            *    10209    10308: gap of unknown length
            *    10309    11734: contig of 1426 bp in length
            *    11735    11834: gap of unknown length
            *    11835    13265: contig of 1431 bp in length
            *    13266    13365: gap of unknown length
            *    13366    14425: contig of 1060 bp in length
            *    14426    14525: gap of unknown length
            *    14526    15609: contig of 1084 bp in length
            *    15610    15709: gap of unknown length
            *    15710    17529: contig of 1820 bp in length
            *    17530    17629: gap of unknown length
            *    17630    20351: contig of 2722 bp in length
            *    20352    20451: gap of unknown length
            *    20452    23753: contig of 3302 bp in length
            *    23754    23853: gap of unknown length
            *    23854    26537: contig of 2684 bp in length
            *    26538    26637: gap of unknown length
            *    26638    31339: contig of 4702 bp in length
            *    31340    31439: gap of unknown length
            *    31440    37041: contig of 5602 bp in length
            *    37042    37141: gap of unknown length
            *    37142    45268: contig of 8127 bp in length
            *    45269    45368: gap of unknown length
            *    45369    52323: contig of 6955 bp in length
            *    52324    52423: gap of unknown length
            *    52424    60239: contig of 7816 bp in length
            *    60240    60339: gap of unknown length
            *    60340    68385: contig of 8046 bp in length
            *    68386    68485: gap of unknown length
            *    68486    78834: contig of 10349 bp in length
            *    78835    78934: gap of unknown length
            *    78935    90422: contig of 11488 bp in length
            *    90423    90522: gap of unknown length
            *    90523   100763: contig of 10241 bp in length
            *   100764   100863: gap of unknown length
            *   100864   111672: contig of 10809 bp in length
            *   111673   111772: gap of unknown length
            *   111773   126855: contig of 15083 bp in length
            *   126856   126955: gap of unknown length
            *   126956   142557: contig of 15602 bp in length
            *   142558   142657: gap of unknown length
            *   142658   161575: contig of 18918 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161575
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-333E1"
BASE COUNT    44288 a  35882 c  35114 g  43686 t   2605 others
ORIGIN


  Query Match             35.5%;  Score 488.8;  DB 2;  Length 161575;
  Best Local Similarity   48.5%;  Pred. No. 1.5e-106;
  Matches  485;  Conservative  202;  Mismatches  280;  Indels   34;  Gaps    2;

Qy    390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
          :||:||:||::: ||:||:||::  ||:||:|  || || ||||| :| :| || |  ||
Db 121612 GGAAGAAATTTCTAAGCAGCAAAACATTCAAGGTGTGACTTGGGTCCTGTTAAATGGATT 121553

Qy    450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
          ::: ||:|| :| ||:|| ||:||:||::: :: ||:||::| || || || ||:|| ||
Db 121552 CAGTTTTATGAGGGAAGCAGAGCATAAAAGTTCAGAAAATTTGCAGCCTGAGAATGTGAT 121493

Qy    510 haaraaraaraayccnttywsngargngnaarttyaarytngcngcngarathtgyathtg 569
          : |:|||:||:||:|| ||::: ||:|  ||:||:|| :  || || ||:||:||:||: |
Db 121492 AGAAAAGAAAAACCCATTTTCTGAGGAGAAATTCAATCAGGCTGCAGAAATTTGCATAAG 121433

Qy    570 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 629
          :||: |: |::| ||:|| ||:|| ||:||:||:|| ||:||: |:::  || | ||:||
Db 121432 TAACAAGAAGCTGAATGTTAATCCGCAAGACAATGGGGAAAATGTCTCCAGGGCATGTCA 121373

Qy    630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 689
          ::  :: :: ||:||: :  |:  ::: :  || |||:| || :| :| ||:|||||:||
Db 121372 GAAGTCTTCACAGCAGCCCCTCCCATCACAGGCCTGGAGGCCTAGGAGAAAATGGTTTTG 121313

Qy    690 yggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtncc 749
          :||  | || || || :: :  ||:||:|| ||:|| :| ||::| || |  ||: | ||
Db 121312 TGGGCCAGGCCCAGGGTCCCAATGCTGTGTCCAGCCTAGGGACTTGGTGCGCTGCATCCC 121253
```

```
Qy     750 ngtnaaywsngcngtngcnwsngarg--------------------------------g 776
Db 121252 AGCCGCTCTAGCCATGGCTGAAAGGGGCCAATGTAGATAGAGCTCAAGCTGTGGCTTCAG 121193

Qy     777 ngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarws 836
Db 121192 AGGGTGCCCTAAGCCTTGGCAGTTTCCACATGGTGTTGAGTCTTTAGGTACACAGAAGTC 121133

Qy     837 nmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggat 896
Db 121132 AAGAATTGAGGTTTGGGAACTTCCACCTGGATTTCAGAAGATGTATGGAAACGCTTGGAT 121073

Qy     897 gccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngt 956
Db 121072 GCCCAGCCAGAAGTTTGCTGCAGGGGCAGGGATCTCATGGAGAACCTCTGCTAGGGCACT 121013

Qy     957 ncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnws 1016
Db 121012 GCAGAAGGGAAATGTGGGGTCGAAGCCACCATATGGAGTCCCTACTCGGGCACCTTCTAG 120953

Qy    1017 nmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayws 1076
Db 120952 TGGAGCTGTGAGAAGAGGGCCAATGACCTCCAGACCCCAGAATGGTAGATCCACCTACAG 120893

Qy    1077 nytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcngg 1136
Db 120892 CTTGCACCGTATGCCTGGAAAAGCTGCAGACACTCAACACCAGCCCGTGAAAGCAGCCAG 120833

Qy    1137 natggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathta 1196
Db 120832 GAGGGAGGCTGTGCCCTACAAAGCCACAGGGGTGGAGTTGCCCAAGACCATGGGAACCCA 120773

Qy    1197 yytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngc 1256
Db 120772 CCTCTTGCATCAGCATGACCTGGATGTGAGACCTGGAGTCAAAGGAGATCATTTTGGAGG 120713

Qy    1257 nytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgyt 1315
Db 120712 TTTAAAATTTGATTGCCTTGCTGGATTTCGGACTTGCATGGGCCCTGTAACCCCTTTGTT 120653

Qy    1316 tyggncarttyttyccnttyggnacngcngtnttyacncar 1356
Db 120652 TTGGCCAATTTCTCCCATTTGGAACAGCTGTATTTACCCAG 120612


RESULT  9
AC087500
LOCUS       AC087500              163674 bp    DNA     linear   HTG 10-JAN-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-333E1 map 17, *** SEQUENCING
            IN PROGRESS ***, 11 unordered pieces.
ACCESSION   AC087500
VERSION     AC087500.4  GI:18104839
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163674)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-333E1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163674)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 10, 2002 this sequence version replaced gi:13560412.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12015
                Center clone name: 333_E_1
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3134: contig of 3134 bp in length
            *     3135  3234: gap of      100 bp
            *     3235    27783: contig of 24549 bp in length
            *    27784  27883: gap of      100 bp
            *    27884    50218: contig of 22335 bp in length
            *    50219  50318: gap of      100 bp
            *    50319    54526: contig of 4208 bp in length
            *    54527  54626: gap of      100 bp
            *    54627    60200: contig of 5574 bp in length
            *    60201  60300: gap of      100 bp
            *    60301    70374: contig of 10074 bp in length
            *    70375  70474: gap of      100 bp
            *    70475    79596: contig of 9122 bp in length
            *    79597  79696: gap of      100 bp
            *    79697    86804: contig of 7108 bp in length
            *    86805  86904: gap of      100 bp
            *    86905   123464: contig of 36560 bp in length
            *   123465  123564: gap of      100 bp
            *   123565   160329: contig of 36765 bp in length
            *   160330  160429: gap of      100 bp
            *   160430   163674: contig of 3245 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163674
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-333E1"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    44782 a  37406 c  35496 g  44882 t   1108 others
ORIGIN


  Query Match          35.5%;  Score 488.8;  DB 2;  Length 163674;
  Best Local Similarity   48.5%;  Pred. No. 1.5e-106;
  Matches  485;  Conservative  202;  Mismatches  280;  Indels   34;  Gaps    2;

Qy     390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
```

```
Db 136131 GGAAGAAATTTCTAAGCAGCAAAACATTCAAGGTGTGACTTGGGTCCTGTTAAATGGATT 136190

Qy    450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
          ::: ||:|| :| ||:|| ||:||:||::: :: ||:||::| || || || ||:|| ||
Db 136191 CAGTTTTATGAGGGAAGCAGAGCATAAAAGTTCAGAAAATTTGCAGCCTGAGAATGTGAT 136250

Qy    510 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 569
          : |:||:||:||:|| ||::: ||:|  ||:||:|| :  || || ||:||:||:||: |
Db 136251 AGAAAAGAAAAACCCATTTTCTGAGGAGAAATTCAATCAGGCTGCAGAAATTTGCATAAG 136310

Qy    570 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 629
          :||: |: |::| ||:|| ||:|| ||:||:||:|| ||:||: |:::  || | ||:||
Db 136311 TAACAAGAAGCTGAATGTTAATCCGCAAGACAATGGGGAAAATGTCTCCAGGGCATGTCA 136370

Qy    630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 689
          ::  :: :: ||:||: :  |:  ::: :  || |||:| || :| :| ||:|||||:||
Db 136371 GAAGTCTTCACAGCAGCCCCTCCCATCACAGGCCTGGAGGCCTAGGAGAAAATGGTTTTG 136430

Qy    690 yggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtncc 749
          :||  | || || || :: :  ||:||:|| ||:|| :| ||::| || |  ||: | ||
Db 136431 TGGGCCAGGCCCAGGGTCCCAATGCTGTGTCCAGCCTAGGGACTTGGTGCGCTGCATCCC 136490

Qy    750 ngtnaaywsngcngtngcnwsngarg--------------------------------g 776
            |   :  ||  | ||      :|                                  |
Db 136491 AGCCGCTCTAGCCATGGCTGAAAGGGGCCAATGTAGATAGAGCTCAAGCTGTGGCTTCAG 136550

Qy    777 ngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarws 836
            :: || ||:|| ||||::| ||    || || ||: |  | ||  |  |:||:::
Db 136551 AGGGTGCCCTAAGCCTTGGCAGTTTCCACATGGTGTTGAGTCTTTAGGTACACAGAAGTC 136610

Qy    837 nmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggat 896
            :| ||:||:|| |||||:|  ||   : | ||:||:||:|| ||:|| ||: | |||||
Db 136611 AAGAATTGAGGTTTGGGAACTTCCACCTGGATTTCAGAAGATGTATGGAAACGCTTGGAT 136670

Qy    897 gccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngt 956
          ||| :| ||:||:||:|| |  || |  || :  :: |||:| || :: || :| |   |
Db 136671 GCCCAGCCAGAAGTTTGCTGCAGGGGCAGGGATCTCATGGAGAACCTCTGCTAGGGCACT 136730

Qy    957 ncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnws 1016
            ||:||:|| ||:|| || | | |:|| ||  |: | || || ::  | || || :: ::
Db 136731 GCAGAAGGGAAATGTGGGGTCGAAGCCACCATATGGAGTCCCTACTCGGGCACCTTCTAG 136790

Qy   1017 nmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayws 1076
            | || || :| :|  : ||    :: :: :| :  ||:|| || :| :: ||  |:::
Db 136791 TGGAGCTGTGAGAAGAGGGCCAATGACCTCCAGACCCCAGAATGGTAGATCCACCTACAG 136850

Qy   1077 nytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcngg 1136
          :| || | : | || | :||: :  | ||:|| ||:  :||:|| || ||:|| ||  |
Db 136851 CTTGCACCGTATGCCTGGAAAAGCTGCAGACACTCAACACCAGCCCGTGAAAGCAGCCAG 136910

Qy   1137 natggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathta 1196
            | |||: : || || ||:||: |    |  |  ||::|  | ||:||  | || | : |
Db 136911 GAGGGAGGCTGTGCCCTACAAAGCCACAGGGGTGGAGTTGCCCAAGACCATGGGAACCCA 136970

Qy   1197 yytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngc 1256
          ::| :| ||:   ||:||::| ||:|| :| | :|| || ||: | ||:||:||:|| |
Db 136971 CCTCTTGCATCAGCATGACCTGGATGTGAGACCTGGAGTCAAAGGAGATCATTTTGGAGG 137030

Qy   1257 nytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgyt 1315
          :| :  ||:||:||:|   | || ||::| || | :||||| || ||  || :| ||:|
Db 137031 TTTAAAATTTGATTGCCTTGCTGGATTTCGGACTTGCATGGGCCCTGTAACCCCTTTGTT 137090

Qy   1316 tyggncarttyttyccnttyggnacngcngtnttyacncar 1356
          |:|| ||:||: |:|| ||:|| || || || ||:|| ||:
Db 137091 TTGGCCAATTTCTCCCATTTGGAACAGCTGTATTTACCCAG 137131


RESULT 10
AC015727/c
LOCUS       AC015727               175639 bp    DNA     linear   HTG 09-SEP-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-420A6 map 17, WORKING DRAFT
            SEQUENCE, 31 unordered pieces.
ACCESSION   AC015727
VERSION     AC015727.4  GI:10045179
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175639)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-420A6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175639)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 9, 2000 this sequence version replaced gi:6958075.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1423
                Center clone name: 420_A_6
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 87% of reads
                Sequencing vector: Plasmid; n/a; 13% of reads
                Chemistry: Dye-primer-amersham; 4% of reads
                Chemistry: Dye-terminator Big Dye; 96% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161512 bases at least Q40
                Consensus quality: 167866 bases at least Q30
                Consensus quality: 170654 bases at least Q20
                Insert size: 170000; agarose-fp
                Insert size: 172639; sum-of-contigs
                Quality coverage: 4.9 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    13947: contig of 13947 bp in length
            *    13948 14047: gap of      100 bp
            *    14048    14259: contig of 212 bp in length
            *    14260 14359: gap of      100 bp
            *    14360    15039: contig of 680 bp in length
            *    15040 15139: gap of      100 bp
            *    15140    15832: contig of 693 bp in length
            *    15833 15932: gap of      100 bp
            *    15933    16689: contig of 757 bp in length
            *    16690 16789: gap of      100 bp
            *    16790    18092: contig of 1303 bp in length
            *    18093 18192: gap of      100 bp
            *    18193    19993: contig of 1801 bp in length
```

```
            *    19994 20093: gap of      100 bp
            *    20094    22205: contig of 2112 bp in length
            *    22206 22305: gap of      100 bp
            *    22306    23915: contig of 1610 bp in length
            *    23916 24015: gap of      100 bp
            *    24016    26544: contig of 2529 bp in length
            *    26545 26644: gap of      100 bp
            *    26645    28851: contig of 2207 bp in length
            *    28852 28951: gap of      100 bp
            *    28952    30454: contig of 1503 bp in length
            *    30455 30554: gap of      100 bp
            *    30555    33015: contig of 2461 bp in length
            *    33016 33115: gap of      100 bp
            *    33116    35727: contig of 2612 bp in length
            *    35728 35827: gap of      100 bp
            *    35828    59255: contig of 23428 bp in length
            *    59256 59355: gap of      100 bp
            *    59356    62539: contig of 3184 bp in length
            *    62540 62639: gap of      100 bp
            *    62640    67930: contig of 5291 bp in length
            *    67931 68030: gap of      100 bp
            *    68031    73174: contig of 5144 bp in length
            *    73175 73274: gap of      100 bp
            *    73275    77427: contig of 4153 bp in length
            *    77428 77527: gap of      100 bp
            *    77528    81788: contig of 4261 bp in length
            *    81789 81888: gap of      100 bp
            *    81889    86379: contig of 4491 bp in length
            *    86380 86479: gap of      100 bp
            *    86480    91547: contig of 5068 bp in length
            *    91548 91647: gap of      100 bp
            *    91648    97783: contig of 6136 bp in length
            *    97784 97883: gap of      100 bp
            *    97884   105190: contig of 7307 bp in length
            *   105191 105290: gap of      100 bp
            *   105291   111855: contig of 6565 bp in length
            *   111856 111955: gap of      100 bp
            *   111956   119820: contig of 7865 bp in length
            *   119821 119920: gap of      100 bp
            *   119921   131299: contig of 11379 bp in length
            *   131300 131399: gap of      100 bp
            *   131400   139546: contig of 8147 bp in length
            *   139547 139646: gap of      100 bp
            *   139647   155538: contig of 15892 bp in length
            *   155539 155638: gap of      100 bp
            *   155639   173481: contig of 17843 bp in length
            *   173482 173581: gap of      100 bp
            *   173582   175639: contig of 2058 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175639
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-420A6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .13947
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    14048. .14259
                     /note="assembly_fragment"
     misc_feature    14360. .15039
                     /note="assembly_fragment"
     misc_feature    15140. .15832
                     /note="assembly_fragment"
     misc_feature    15933. .16689
                     /note="assembly_fragment"
     misc_feature    16790. .18092
                     /note="assembly_fragment"
     misc_feature    18193. .19993
                     /note="assembly_fragment"
     misc_feature    20094. .22205
                     /note="assembly_fragment"
     misc_feature    22306. .23915
                     /note="assembly_fragment"
     misc_feature    24016. .26544
                     /note="assembly_fragment"
     misc_feature    26645. .28851
                     /note="assembly_fragment"
     misc_feature    28952. .30454
                     /note="assembly_fragment"
     misc_feature    30555. .33015
                     /note="assembly_fragment"
     misc_feature    33116. .35727
                     /note="assembly_fragment"
     misc_feature    35828. .59255
                     /note="assembly_fragment"
     misc_feature    59356. .62539
                     /note="assembly_fragment"
     misc_feature    62640. .67930
                     /note="assembly_fragment"
     misc_feature    68031. .73174
                     /note="assembly_fragment"
     misc_feature    73275. .77427
                     /note="assembly_fragment"
     misc_feature    77528. .81788
                     /note="assembly_fragment"
     misc_feature    81889. .86379
                     /note="assembly_fragment"
     misc_feature    86480. .91547
                     /note="assembly_fragment"
     misc_feature    91648. .97783
                     /note="assembly_fragment"
     misc_feature    97884. .105190
                     /note="assembly_fragment"
     misc_feature    105291. .111855
                     /note="assembly_fragment"
     misc_feature    111956. .119820
                     /note="assembly_fragment"
     misc_feature    119921. .131299
                     /note="assembly_fragment"
     misc_feature    131400. .139546
                     /note="assembly_fragment"
     misc_feature    139647. .155538
                     /note="assembly_fragment"
     misc_feature    155639. .173481
                     /note="assembly_fragment"
     misc_feature    173582. .175639
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    48575 a  37483 c  38030 g  48544 t   3007 others

  Query Match            35.5%;  Score 488.8;  DB 2;  Length 175639;
  Best Local Similarity  48.5%;  Pred. No. 1.5e-106;
  Matches  485;  Conservative  202;  Mismatches  280;  Indels   34;  Gaps    2;

Qy     390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
           :||:||:||::: ||:||:||::  ||:||:|  || || ||||| :| :| || |  ||
Db    2929 GGAAGAAATTTCTAAGCAGCAAAACATTCAAGGTGTGACTTGGGTCCTGTTAAATGGATT 2870

Qy     450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
           ::: ||:|| :| ||:|| ||:||:||::: :: ||:||::| || || || ||:|| ||
Db    2869 CAGTTTATGAGGGAAGCAGAGCATAAAAGTTCAGAAAATTTGCAGCCTGAGAATGTGAT 2810

Qy     510 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 569
           : |:||:||:||:|| ||::: ||:|  ||:||:|| :  || || ||:||:||:||: |
Db    2809 AGAAAAGAAAAACCCATTTTCTGAGGAGAAATTCAATCAGGCTGCAGAAATTTGCATAAG 2750

Qy     570 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 629
           :||: |: |::| ||:|| ||:|| ||:||:||:|| ||:||: |:::  || | ||:||
Db    2749 TAACAAGAAGCTGAATGTTAATCCACAAGACAATGGGGAAAATGTCTCCAGGGCATGTCA 2690

Qy     630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 689
           ::  :: :: ||:||: :  |:  ::: :  || |||:| || :| :| ||:||||||
```

```
Db   2689 GAAGTCTTCACAGCAGCCCCTCCCATCACAGGCCTGGAGGCCTAGGAGAAAATGGTTTTG 2630

Qy    690 yggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtncc 749
          :||  | || || || :: :  ||:||:|| ||:|| :| ||::| || |  ||: | ||
Db   2629 TGGGCCAGGCCCAGGGTCCCAATGCTGTGTCCAGCCTAGGGACTTGGTGCGCTGCATCCC 2570

Qy    750 ngtnaaywsngcngtngcnwsngarg--------------------------------g 776
           |    :   ||  | ||      :|                                 |
Db   2569 AGCCGCTCTAGCCATGGCTGAAAGGGGCCAATGTAGATAGAGCTCAAGCTGTGGCTTCAG 2510

Qy    777 ngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarws 836
           |  :: || ||:|| |||||::| ||    || || ||: |  | ||  |  |:||:::
Db   2509 AGGGTGCCCTAAGCCTTGGCAGTTTCCACATGGTGTTGAGTCTTTAGGTACACAGAAGTC 2450

Qy    837 nmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggat 896
           :| ||:||:|| |||||:|  ||   : | ||:||:||:|| ||:|| ||: | |||||
Db   2449 AAGAATTGAGGTTTGGGAACTTCCACCTGGATTTCAGAAGATGTATGGAAACGCTTGGAT 2390

Qy    897 gccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngt 956
          ||| :| ||:||:||:|| |  || |  || :  :: |||:| || :: || :| |   |
Db   2389 GCCCAGCCAGAAGTTTGCTGCAGGGGCAGGGATCTCATGGAGAACCTCTGCTAGGGCACT 2330

Qy    957 ncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnws 1016
           ||:||:|| ||:|| || | | |:|| ||  |: | || || ::  | || || :: ::
Db   2329 GCAGAAGGGAAATGTGGGGTCGAAGCCACCATATGGAGTCCCTACTCGGGCACCTTCTAG 2270

Qy   1017 nmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayws 1076
            | || || :| :|  : ||    :: :: :| :  ||:|| || :| :: ||  |:::
Db   2269 TGGAGCTGTGAGAAGAGGGCCAATGACCTCCAGACCCCAGAATGGTAGATCCACCTACAG 2210

Qy   1077 nytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcngg 1136
           :| || | : | || | :||: :  | ||:|| ||:  :||:|| || ||:|| ||  |
Db   2209 CTTGCACCGTATGCCTGGAAAAGCTGCAGACACTCAACACCAGCCCGTGAAAGCAGCCAG 2150

Qy   1137 natggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathta 1196
           | |||: : || || ||:||: |    |  |  ||::|  | ||:||  | || | : |
Db   2149 GAGGGAGGCTGTGCCCTACAAAGCCACAGGGGTGGAGTTGCCCAAGACCATGGGAACCCA 2090

Qy   1197 yytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngc 1256
          ::| :| ||:   ||:||::| ||:|| :| | :|| || ||: | ||:|||:||:|| |
Db   2089 CCTCTTGCATCAGCATGACCTGGATGTGAGACCTGGAGTCAAAGGAGATCATTTTGGAGG 2030

Qy   1257 nytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgyt 1315
           :| :  ||:||:||:|   | || ||::| || | :||||| || ||  || :| ||:|
Db   2029 TTTAAAATTTGATTGCCTTGCTGGATTTCGGACTTGCATGGGCCCTGTAACCCCTTTGTT 1970

Qy   1316 tyggncarttyttyccnttyggnacngcngtnttyacncar 1356
          |:|| ||:||: |:|| ||:|| || || || ||:|| ||:
Db   1969 TTGGCCAATTTCTCCCATTTGGAACAGCTGTATTTACCCAG 1929
```

```
RESULT  11
AL391823/c
LOCUS       AL391823              181842 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-319F11, *** SEQUENCING IN
            PROGRESS ***, 12 unordered pieces.
ACCESSION   AL391823
VERSION     AL391823.9  GI:11995113
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181842)
  AUTHORS   Mclay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 28, 2000 this sequence version replaced gi:11990058.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA319F11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 4% of reads
            Sequencing vector: plasmid; L08752; 95% of reads
            Chemistry: Dye-terminator ET-amersham; 3% of reads Chemistry:
            Dye-terminator Big Dye; 96% of reads
            Consensus quality: 174456 bases at least Q40
            Consensus quality: 177550 bases at least Q30
            Consensus quality: 179392 bases at least Q20
            Insert size: 180742; sum-of-contigs
            Insert size: 162786; 7.7% error; agarose-fp
            Quality coverage: 4.40x in Q20 bases; sum-of-contigs Quality
            coverage: 4.89x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     8896: contig of 8896 bp in length
            *     8897 8996: gap of      100 bp
            *     8997    22635: contig of 13639 bp in length
            *    22636 22735: gap of      100 bp
            *    22736    55625: contig of 32890 bp in length
            *    55626 55725: gap of      100 bp
            *    55726    65440: contig of 9715 bp in length
            *    65441 65540: gap of      100 bp
            *    65541    88588: contig of 23048 bp in length
            *    88589 88688: gap of      100 bp
            *    88689    94602: contig of 5914 bp in length
            *    94603 94702: gap of      100 bp
            *    94703   105410: contig of 10708 bp in length
            *   105411 105510: gap of      100 bp
            *   105511   107644: contig of 2134 bp in length
            *   107645 107744: gap of      100 bp
            *   107745   126586: contig of 18842 bp in length
            *   126587 126686: gap of      100 bp
            *   126687   155610: contig of 28924 bp in length
            *   155611 155710: gap of      100 bp
            *   155711   173103: contig of 17393 bp in length
            *   173104 173203: gap of      100 bp
            *   173204   181842: contig of 8639 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181842
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-319F11"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .8896
                     /note="assembly_fragment:02130
                     clone_end:SP6
                     vector_side:left"
     misc_feature    8997. .22635
                     /note="assembly_fragment:01909
                     fragment_chain:1"
     misc_feature    22736. .55625
                     /note="assembly_fragment:00961
                     fragment_chain:1"
     misc_feature    55726. .65440
                     /note="assembly_fragment:00129
                     fragment_chain:1"
     misc_feature    65541. .88588
                     /note="assembly_fragment:00944
                     fragment_chain:1"
     misc_feature    88689. .94602
```

```
                     /note="assembly_fragment:01651
                     fragment_chain:1"
     misc_feature    94703. .105410
                     /note="assembly_fragment:01169
                     fragment_chain:1"
     misc_feature    105511. .107644
                     /note="assembly_fragment:00035
                     fragment_chain:1"
     misc_feature    107745. .126586
                     /note="assembly_fragment:01445
                     fragment_chain:1"
     misc_feature    126687. .155610
                     /note="assembly_fragment:00583
                     fragment_chain:1"
     misc_feature    155711. .173103
                     /note="assembly_fragment:00742
                     fragment_chain:1"
     misc_feature    173204. .181842
                     /note="assembly_fragment:01461"
BASE COUNT    53809 a  39941 c  37882 g  49105 t   1105 others
ORIGIN


  Query Match             35.5%;  Score 488.6;  DB 2;  Length 181842;
  Best Local Similarity   47.5%;  Pred. No. 1.7e-106;
  Matches  486;  Conservative  212;  Mismatches  290;  Indels   36;  Gaps    2;

Qy     389 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnt 448
           |:||:| : |: : ||:||:||::: ||:  :|| || || | ||| :| :| ||:|| |
Db   64835 AAGAAGGATTTCCTAAGCAGCAAAGCATTTGAGATGTGACTTTGGTGCTGTTAAAGGCAT 64776

Qy     449 tywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtna 508
           |::: ||:||::| ||:|| || ||:||::  :  ||:||::| || || ||: |:|  |
Db   64775 TCAGTTTTATAAGGGAAGCAGATCATAAAAATTTGGAAAATTTGCAGCCTGACTATGCAA 64716

Qy     509 thaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyatht 568
           |: |:||:||:||:|| ||::: | :|  ||:||:||::     || ||:||:||: |:
Db   64715 TAGAAAAGAAAAACCCATTTTCTGGGGAGAAATTCAAGCCTATTGCAGAAATTTGCCTAA 64656

Qy     569 gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 628
           |:| : |:||::  ||:|| ||:|| ||:||: |: | | :||: |::   || | ||:|
Db   64655 GTAGCAAGGAGCCTAATGTTAATCCCCAAGACCATCGGGGAAATGTCTACAGGCCATGTC 64596

Qy     629 armgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naartggt 685
           |: | :: ::   :| : :  |:  ::: :  || || :| || :| :|    ||:||||
Db   64595 AGGGAACTTCATGGCTGCCCCTCCCATCACAAGCCTGCAGGCCCAGGAGTAAAAAGTGGT 64536

Qy     686 tytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgyg 745
           |:  :||  | || || || :  :| ||: |:|| ||:|  :| ||::| || || ||:|
Db   64535 TTCATGGGCCAGGCCCAGGGTTCCTGTGCCGTGTGCAGCACAGGGACTTGGTGCCCTGTG 64476

Qy     746 tnccng----------------------------------tnaaywsngcngtngcnwsng 772
           | || |                                  ::: || || || :: |
Db   64475 TCCCAGCTGCTCCAGCTATGGCTGAAAGGGGCCAAGGTAGAGGTTGGGCTGTGGCTTCAG 64416

Qy     773 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 832
           |:|| || :: || ||:|| |||||::| || . : |  || ||:|| || || ||  |:|
Db   64415 AGGGTGCAAGACCCAAGCCTTGGCAGCTTCCACGTGATGTTGAGCCTGTGGGTGCACAGA 64356

Qy     833 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 892
           |::: :| | :||:|| |||||:|| |    ::| ||:||:||:|| ||:|| ||: | |
Db   64355 AGTCAAGAACTGAGGTTTGGGAACCTCTGCCTAGATTTCAGAAGATGTATGGAAACGCCT 64296

Qy     893 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 952
           ||||||| :| ||:||:||:|| |  || || ||  : :: |||:| || :: || :| |
Db   64295 GGATGCCCAGGCAAAAGTTTGCTGCAGGTGTGGGACCCTCATGGAGAACCTCTGCTAGGG 64236

Qy     953 tngtncaraarggnaaygtnggntgggarccncccncaymgngtnccnwsnggngcnccnw 1012
             || ||:||:|  ||:|| || |  ||:|| |  ||::| || || :: || || ||
Db   64235 CAGTGCAGAAGGAAAATGTGGGGTCAGAGCCCCACCACAGAGTCCCTACTGGGGCACCAC 64176

Qy    1013 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1072
           : ::  | || || :| :|  : || |  :: :: :| :  ||:|| |  :| :: || |

Db   64175 CTAGTGGAGCTGTGAGAAGGGGGCCACTGTCCTCCAGACCCCAGAATGATAGATCCACTG 64116

Qy    1073 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1132
           |::: :| || ||:   || | :||: : || ||:|| ||: |:||:||  | ||:|| |
Db   64115 ACAGCTTGCACCATTCACCTGGAAAAGCCACAGACACTCAACGCCAGCCCATGAAAGCAG 64056

Qy    1133 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1192
           |  | | |||: : || || | :||: |    |  |  ||::|  | ||:||  | || |
Db   64055 CCAGGAGGGAGGCTGTACCCTGCAAAGCCACAGGGGTGGAACTGCCCAAGACTATGGGAA 63996

Qy    1193 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1252
            :||::| :| ||:   ||:||::| ||:|| :| | :|| || ||: | ||:||:||:|
Db   63995 CCTACCTCTTGCATCAGCATGACCTGGATGTGAGACCTGGTGTCAAAGGAGATCATTTTG 63936

Qy    1253 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytnt 1312
             || :| :  ||:||:||:|  || || ||::  || | :||||| || || || :| |
Db   63935 AAGCTTTAAAATTTGACTGCCTCACTGGATTTCAGACTTGCATGGGCCCTGTACCACTTT 63876

Qy    1313 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1372
           |:||:|| ||:||: |:|| ||:|| |  ||  | ||:|| ||:| ::| ||::  ||:|
Db   63875 GTTTTGGCCAATTTCTCCCATTTGGGAGGGCTATATTTACCCAGTACTTGTACCCCCATT 63816

Qy    1373 gyat 1376
           |:||
Db   63815 GTAT 63812


RESULT  12
AL353634
LOCUS       AL353634               77405 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome X clone RP11-435A2 map q21.33-22.3, ***
            SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION   AL353634
VERSION     AL353634.4  GI:9863649
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 77405)
  AUTHORS   Mclay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 21, 2000 this sequence version replaced gi:9213111.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA435A2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 13% of reads Chemistry:
            Dye-terminator Big Dye; 86% of reads
            Consensus quality: 76604 bases at least Q40
            Consensus quality: 77098 bases at least Q30
            Consensus quality: 77320 bases at least Q20
            Insert size: 77405; sum-of-contigs
            Insert size: 85298; 1.5% error; agarose-fp
            Quality coverage: 4.54x in Q20 bases; sum-of-contigs Quality
            coverage: 4.30x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .77405
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q21.33-22.3"
                     /clone="RP11-435A2"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .77405
                     /note="assembly_fragment:00554"
BASE COUNT     25072 a  14551 c  15133 g  22649 t
ORIGIN
```

Query Match 35.5%; Score 488.2; DB 2; Length 77405;
Best Local Similarity 47.0%; Pred. No. 1.6e-106;
Matches 508; Conservative 225; Mismatches 310; Indels 37; Gaps 4;

```
Qy     334 ytncaycaytgyaargtnaayathtggytnatgmgnaarcarathytngcnaayaargar 393
Db    8637 CTGCCCTAGAGATCTGTGGAATTTGACCTTGAGAGAGATGATTTAGGGTATCAGGTGGAA 8696

Qy     394 garathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnttywsn 453
Db    8697 GAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGTGTCCTGTTAAAGGCACTCAGT 8756

Qy     454 ttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnathaar 513
Db    8757 TTTGCAAGGGAAGCAGAACATAAAAGTTCAGAAAATTTTCAGCCTGACAATGTGATAGAA 8816

Qy     514 aaraaraaycc-nttywsngarggnaarttyaarytngcngcngarathtgyathtgyaa 572
Db    8817 AAGAAAACCCCATTTTTCTGATGATAAATTGAAGCCAGCTGCAGAAATTTGCATAAGTAA 8876

Qy     573 ygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycarmg 632
Db    8877 CGAGGAGCCAAATGTTAATCACTAAGACCATGGGGAAAATATCTCCATGGCTTGTCAGAG 8936

Qy     633 nwsnwsncarcarwsnathaarwsnytng--cntggmgnccnmgnmgnaartggttytgy 690
Db    8937 GTCTTCACGGCAGCCCCTCACATCACTGGCCCCCAGAGGCCTAGGAGAAAATGGTTTTAT 8996

Qy     691 ggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtnccn 750
Db    8997 GGGTCGGGCCCAGGGTCCCCATGCTATGTGCAGTCTAGGGACTTGGTGCCCCGTGTCCCA 9056

Qy     751 gtnaa---------------------------------ywsngcngtngcnwsngarggn 777
Db    9057 GCCACTCCAGCCATAACTGAAAGGGGCCAAGGTATAGCTTGGGCTGTTGCTTCAAAGGAT 9116

Qy     778 gcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsn 837
Db    9117 GGAAGCCCCAGGCCTTGGCAGCTTCCATGTGGTGTTGAGCCTGCAGGTGCACTGAAGTCA 9176

Qy     838 mgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggatg 897
Db    9177 AGAATTGAGGTTTGGGAACCTCCACCTAGATTTCAGAAGATGTATGGAAATGCCTGGATG 9236

Qy     898 ccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtn 957
Db    9237 CCCAGGCAAAAGTTTGCTGCAGGGGCGAGGACCTCATGGAGAACCTCTGTTAGGGTGGTG 9296

Qy     958 caraarggnaaygtnggntgggarccncccncaymgngtnccnwsnggngcnccnwsnwsn 1017
Db    9297 TGGAAGGGAAAAGTGGGGTCATAGCCGCCACACAGAGTCCCTACCAGGGGACCACCTAGT 9356

Qy    1018 mgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsn 1077
Db    9357 GGAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAATGGTAGATCCACCGACAGC 9416

Qy    1078 ytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggn 1137
Db    9417 CTGCACCGTGTGCCTGGAAAAGCCGCAGACACTCAATGCCAGCCTGTGAAAGCAGCAGGG 9476

Qy    1138 atggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtay 1197
```

```
Db    9477 AGGGAGGCTGTACCCTGCAAAGCCACAGGGGCAGAGCTGCCCAAGATCATGGGAACCAAT 9536

Qy    1198 ytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcn 1257
Db    9537 CTCTTGCATCAGCATGACCTGGATATGAGACATGGAGTCAAAGGAAATCATTTTGGAGCT 9596

Qy    1258 ytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgytt 1316
Db    9597 TTAAGATTTGACTGTCCCACTGGATCTCAGACTTCCATGGGGCCTGTAGCCTCTTTGTTT 9656

Qy    1317 yggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytgyat 1376
Db    9657 TGGCCAATATCTCCCATTTGGAATGGCTGTATTTATCCAATGCCTGTACCCCCATTGTAT 9716
```

```
RESULT  13
HSJ633H17/c
LOCUS       HSJ633H17             124531 bp    DNA     linear   PRI 07-FEB-2000
DEFINITION  Human DNA sequence from clone 633H17 on chromosome lp31.2-32.2.
            Contains a pseudogene similar to part of MTCO1 (Cytochrome C
            oxidase 1), MTCO2 (cytochrome c oxidase II), MTATP8 (ATP synthase
            8) and GOT2 glutamic-oxaloacetic transaminase 2, mitochondrial
            (aspartate aminotransferase 2) pseudogenes, ESTs and GSSs, complete
            sequence.
ACCESSION   AL049710
VERSION     AL049710.18  GI:5650682
KEYWORDS    HTG; aspartate aminotransferase; ATP synthase; Cytochrome C
            oxidase; glutamic-oxaloacetic transaminase; GOT2; MTATP8; MTCO1;
            MTCO2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124531)
  AUTHORS   Moore,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 29, 1999 this sequence version replaced gi:5566565.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence is the entire insert of clone 633H17. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre Chromosome 1
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            633H17 is from the library RPCI-4 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2 The
            following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep.
FEATURES              Location/Qualifiers
     source           1. .124531
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="1"
                      /map="p31.2-32.2"
                      /clone_lib="RPCI-4"
                      /clone="633H17"
```

```
repeat_region   1. .234
                /note="A1uSx repeat: matches 59. .292 of consensus"
misc_feature    complement(235. .546)
                /note="match: GSS: Em:AQ055811"
repeat_region   559. .760
                /note="MIR repeat: matches 74. .262 of consensus"
repeat_region   1158. .2018
                /note="L1MA4 repeat: matches 5282. .6145 of consensus"
repeat_region   2026. .2590
                /note="HERVL repeat: matches 4676. .5241 of consensus"
repeat_region   2879. .9441
                /note="L1PA2 repeat: matches 1. .6146 of consensus"
repeat_region   10000. .10124
                /note="MIR repeat: matches 34. .154 of consensus"
repeat_region   10623. .10689
                /note="MIR repeat: matches 36. .102 of consensus"
repeat_region   10690. .11014
                /note="L2 repeat: matches 2389. .2710 of consensus"
repeat_region   11136. .11262
                /note="L2 repeat: matches 2349. .2473 of consensus"
repeat_region   13420. .13522
                /note="MIR repeat: matches 114. .219 of consensus"
repeat_region   13602. .13725
                /note="MIR repeat: matches 130. .262 of consensus"
repeat_region   13739. .13987
                /note="L2 repeat: matches 2397. .2666 of consensus"
repeat_region   14368. .14486
                /note="L2 repeat: matches 2585. .2705 of consensus"
repeat_region   14656. .14764
                /note="MIR repeat: matches 124. .252 of consensus"
misc_feature    14765. .14965
                /note="match: GSS: Em:B30406"
repeat_region   14966. .15150
                /note="MIR repeat: matches 13. .217 of consensus"
repeat_region   15157. .15818
                /note="MER39b repeat: matches 14. .579 of consensus"
misc_feature    16403. .16945
                /note="match: GSS: Em:AQ454645"
misc_feature    16410. .16933
                /note="match: GSS: Em:AQ357205"
misc_feature    16410. .17037
                /note="match: GSS: Em:AQ309712"
misc_feature    complement(16459. .16861)
                /note="match: GSS: Em:AQ197233"
repeat_region   16531. .16576
                /note="23 copies 2 mer tt 74% conserved"
repeat_region   18053. .18154
                /note="MIR repeat: matches 7. .107 of consensus"
repeat_region   18377. .18492
                /note="MER39b repeat: matches 476. .574 of consensus"
misc_feature    18525. .18954
                /note="match: GSS: Em:AQ210513"
repeat_region   19026. .19126
                /note="MIR repeat: matches 10. .114 of consensus"
repeat_region   19866. .20007
                /note="MIR repeat: matches 28. .169 of consensus"
repeat_region   20273. .20334
                /note="LTR16C repeat: matches 90. .152 of consensus"
repeat_region   20403. .20505
                /note="MER81 repeat: matches 1. .110 of consensus"
repeat_region   20508. .20690
                /note="LTR16C repeat: matches 190. .376 of consensus"
repeat_region   20692. .20725
                /note="L2 repeat: matches 2660. .2693 of consensus"
repeat_region   20828. .20863
                /note="18 copies 2 mer ac 94% conserved"
repeat_region   21172. .21304
                /note="MER5A repeat: matches 40. .167 of consensus"
repeat_region   21201. .21331
                /note="MER5A repeat: matches 56. .188 of consensus"
repeat_region   21466. .21528
                /note="MER54A repeat: matches 214. .285 of consensus"
repeat_region   21952. .22077
                /note="L2 repeat: matches 2578. .2707 of consensus"
repeat_region   22083. .22259
                /note="FRAM repeat: matches -2. .175 of consensus"
repeat_region   22280. .22393
                /note="L2 repeat: matches 2541. .2645 of consensus"
repeat_region   22394. .22748
                /note="MLT1B repeat: matches 1. .390 of consensus"
repeat_region   22749. .22846
                /note="L2 repeat: matches 2645. .2745 of consensus"
repeat_region   22864. .23010
                /note="L2 repeat: matches 2329. .2429 of consensus"
repeat_region   23011. .23310
                /note="A1uJb repeat: matches 1. .299 of consensus"
repeat_region   23311. .23391
                /note="L2 repeat: matches 2429. .2521 of consensus"
repeat_region   23433. .23730
                /note="MSTD repeat: matches 66. .394 of consensus"
repeat_region   23731. .25397
                /note="L1PB1 repeat: matches 4442. .6123 of consensus"
repeat_region   25477. .25556
                /note="MSTD repeat: matches 1. .85 of consensus"
repeat_region   25557. .25651
                /note="L2 repeat: matches 2585. .2692 of consensus"
repeat_region   25745. .25785
                /note="L2 repeat: matches 2702. .2742 of consensus"
repeat_region   26898. .27373
                /note="MER31A repeat: matches 1. .661 of consensus"
repeat_region   27378. .27483
                /note="MER93 repeat: matches 292. .397 of consensus"
misc_feature    27484. .27868
                /note="match: GSS: Em:AQ127976"
repeat_region   27537. .27688
                /note="MER93 repeat: matches 1. .188 of consensus"
repeat_region   27912. .28051
                /note="L1PA13 repeat: matches 6011. .6152 of consensus"
repeat_region   29900. .30121
                /note="HERVL repeat: matches 3020. .3235 of consensus"
repeat_region   30122. .30487
                /note="MLT1A2 repeat: matches 1. .374 of consensus"
repeat_region   30488. .30578
                /note="HERVL repeat: matches 2933. .3020 of consensus"
repeat_region   30579. .30796
                /note="MER58A repeat: matches 1. .221 of consensus"
repeat_region   30797. .31887
                /note="HERVL repeat: matches 1792. .2933 of consensus"
misc_feature    complement(31888. .32090)
                /note="match: GSS: Em:B58153; match: GSS: Em:B66544;
                match: GSS: Em:AQ759038"
misc_feature    32922. .33453
                /note="match: GSS: Em:AQ751315"
repeat_region   33175. .33412
                /note="L2 repeat: matches 2473. .2708 of consensus"
repeat_region   33454. .33885
                /note="MLT1C repeat: matches 6. .466 of consensus"
repeat_region   33896. .33946
                /note="MIR repeat: matches 162. .215 of consensus"
repeat_region   33947. .34302
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   34303. .34334
                /note="MIR repeat: matches 129. .162 of consensus"
repeat_region   34335. .34703
                /note="THE1C repeat: matches 1. .369 of consensus"
repeat_region   34704. .34830
                /note="MIR repeat: matches 20. .129 of consensus"
repeat_region   34831. .35552
                /note="L1MC2 repeat: matches 5573. .6325 of consensus"
repeat_region   36612. .36657
                /note="L2 repeat: matches 2675. .2723 of consensus"
repeat_region   36936. .37004
                /note="L2 repeat: matches 2638. .2707 of consensus"
repeat_region   37383. .37648

Query Match             35.5%;  Score 488.2;  DB 9;  Length 124531;
```

```
  Best Local Similarity   47.4%;  Pred. No. 1.9e-106;
  Matches  485;  Conservative  213;  Mismatches  290;  Indels   36;  Gaps    2;

Qy     389 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnt 448
           |:||:| : |: : ||:||:||::: ||:  :|| || || | ||| :| :| ||:|| |
Db   96092 AAGAAGGATTTCCTAAGCAGCAAAGCATTTGAGATGTGACTTTGGTGCTGTTAAAGGCAT 96033

Qy     449 tywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtna 508
           |::: ||:||::| ||:|| ||:||:||::  :  ||:||::| || || ||: |:|  |
Db   96032 TCAGTTTTATAAGGGAAGCAGAGCATAAAAATTTGGAAAATTTGCAGCCTGACTATGCAA 95973

Qy     509 thaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyatht 568
           |: |:||:||:||:|| ||::: | :|  ||: |:||:     || ||:||:||: |:
Db   95972 TAGAAAAGAAAAACCCATTTTCTGGGGAGAAACTCAAGCCTATTGCAGAAATTTGCCTAA 95913

Qy     569 gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 628
           |:| : |:||::  ||:|| ||:|| ||:||: |: | | :||: |::   || | ||:|
Db   95912 GTAGCAAGGAGCCTAATGTTAATCCCCAAGACCATCGGGGAAATGTCTACAGGCCATGTC 95853

Qy     629 armgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naartggt 685
           |: | :: ::   :| : :  |:  ::: :  || || :| || :| :|    ||:||||
Db   95852 AGGGAACTTCATGGCTGCCCCTCCCATCACAAGCCTGCAGGCCCAGGAGTAAAAAGTGGT 95793

Qy     686 tytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgyg 745
           |:  :||  | || || || :  :| ||: |:|| ||:|  :| ||::| || || ||:|
Db   95792 TTCATGGGCCAGGCCCAGGGTTCCTGTGCCGTGTGCAGCACAGGGACTTGGTGCCCTGTG 95733

Qy     746 tnccng--------------------------------tnaaywsngcngtngcnwsng 772
           | || |                                ::: || || || :: | |
Db   95732 TCCCAGCTGCTCCAGCTATGGCTGAAAGGGGCCAAGGTAGAGGTTGGGCTGTGGCTTCAG 95673

Qy     773 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 832
           |:|| || :: || ||:|| |||||::| ||  : |  || ||:|| || || ||  |:|
Db   95672 AGGGTGCAAGACCCAAGCCTTGGCAGCTTCCACGTGATGTTGAGCCTGTGGGTGCACAGA 95613

Qy     833 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 892
           |::: :| | :||:|| |||||:|| |    ::| ||:||:||:|| ||:|| ||: | |
Db   95612 AGTCAAGAACTGAGGTTTGGGAACCTCTGCCTAGATTTCAGAAGATGTATGGAAACGCCT 95553

Qy     893 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 952
           ||||||| :| ||:||:||:|| |  || || ||  : :: |||:| || :: || :| |
Db   95552 GGATGCCCAGGCAAAAGTTTGCTGCAGGTGTGGGACCCTCATGGAGAACCTCTGCTAGGG 95493

Qy     953 tngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnw 1012
             || ||:||:|  ||:|| || |  ||:|| |  ||::| || || :: || || ||
Db   95492 CAGTGCAGAAGGAAAATGTGGGGTCAGAGCCCCACCACAGAGTCCCTACTGGGGCACCAC 95433

Qy    1013 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1072
           : ::  | || || :| :|  : || |  :: :: :| :  ||:|| |  :| :: || |
Db   95432 CTAGTGGAGCTGTGAGAAGGGGGCCACTGTCCTCCAGACCCCAGAATGATAGATCCACTG 95373

Qy    1073 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1132
           |::: :| || ||:   || | :||: : || ||:|| ||: |:||:||  | ||:|| |
Db   95372 ACAGCTTGCACCATTCACCTGGAAAAGCCACAGACACTCAACGCCAGCCCATGAAAGCAG 95313

Qy    1133 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1192
           |  | | |||: : || || | :||: |    |  |  ||::|  | ||:||  | || |
Db   95312 CCAGGAGGGAGGCTGTACCCTGCAAAGCCACAGGGGTGGAACTGCCCAAGACTATGGGAA 95253

Qy    1193 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1252
            :||::| :| ||:   ||:||::| ||:|| :| | :|| || ||: | ||:||:||:|
Db   95252 CCTACCTCTTGCATCAGCATGACCTGGATGTGAGACCTGGTGTCAAAGGAGATCATTTTG 95193

Qy    1253 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytnt 1312
             || :| :  ||:||:||:|  || || ||::  || | :|||| || || || :| |
Db   95192 AAGCTTTAAAATTTGACTGCCTCACTGGATTTCAGACTTGCATGGGCCCTGTACCACTTT 95133

Qy    1313 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1372
           |:||:|| ||:||: |:|| ||:|| |  ||  | ||:|| ||:| ::| ||::  ||:|
Db   95132 GTTTTGGCCAATTTCTCCCATTTGGGAGGGCTATATTTACCCAATACTTGTACCCCCATT 95073

Qy    1373 gyat 1376
           |:||
```

```
Db  95072 GTAT 95069


RESULT  14
HSU212C1
LOCUS       HSU212C1                40714 bp    DNA     linear   PRI 23-NOV-1999
DEFINITION  Human DNA sequence from cosmid U212C1, between markers DXS366 and
            DXS87 on chromosome X.
ACCESSION   Z69722
VERSION     z69722.1  GI:1204119
KEYWORDS    X.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40714)
  AUTHORS   Odell,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1995) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, UK. E-mail enquires: humpub@sanger.ac.uk
COMMENT     IMPORTANT:  This sequence is the entire insert of clone U212C1. The
            true left end of clone U212C1 is at 1 in this sequence. The true
            right end of clone U212C1 is at 40714.
            U212C1 is from the human chromosome X-specific cosmid library.
FEATURES             Location/Qualifiers
     source          1. .40714
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="LL0XNC01-212C1"
                     /clone_lib="LL0XNC01"
     repeat_region   10. .233
                     /note="L1 element fragment"
     repeat_region   388. .758
                     /note="L1 element fragment"
     repeat_region   932. .1108
                     /note="L1 element fragment"
     repeat_region   2571. .2617
                     /note="MER43 element fragment"
     repeat_region   2613. .5203
                     /note="L1 element fragment"
     repeat_region   5204. .5229
                     /note="13 copies of 2 mer 100 % conserved"
     repeat_region   5230. .5427
                     /note="MER43 element fragment"
     repeat_region   8954. .9088
                     /note="L1 element fragment"
     repeat_region   12171. .13985
                     /note="L1 element fragment"
     repeat_region   14179. .14612
                     /note="L1 element fragment"
     repeat_region   14662. .14780
                     /note="L1 element fragment"
     repeat_region   14824. .15044
                     /note="L1 element fragment"
     repeat_region   15234. .15369
                     /note="L1 element fragment"
     repeat_region   15421. .15849
                     /note="L1 element fragment"
     repeat_region   16026. .16155
                     /note="MSTA element fragment"
     repeat_region   16073. .16149
                     /note="THE1B element fragment"
     repeat_region   16204. .16495
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   16532. .16609
                     /note="THE1B element fragment"
     repeat_region   16624. .16722
                     /note="MSTC element fragment"
     repeat_region   16626. .16737
                     /note="MSTA element fragment"
     repeat_region   16651. .16712
```

```
                     /note="THE1B element fragment"
     repeat_region   17060. .17653
                     /note="L1 element fragment"
     repeat_region   19720. .19797
                     /note="MER21B element fragment"
     repeat_region   19877. .19915
                     /note="MER21B element fragment"
     repeat_region   20007. .20256
                     /note="MER21B element fragment"
     repeat_region   20518. .20734
                     /note="MER30 element fragment"
     repeat_region   22661. .22730
                     /note="MLT1B element fragment"
     repeat_region   22740. .22814
                     /note="MLT1A element fragment"
     repeat_region   22917. .22974
                     /note="MLT1A element fragment"
     repeat_region   23342. .23501
                     /note="THE1B element fragment"
     repeat_region   23550. .23697
                     /note="THE1B element fragment"
     repeat_region   23725. .24279
                     /note="THR element fragment"
     repeat_region   24395. .24593
                     /note="THR element fragment"
     repeat_region   24643. .25270
                     /note="THR element fragment"
     repeat_region   25277. .25463
                     /note="THE1B element fragment"
     repeat_region   25528. .25627
                     /note="THE1B element fragment"
     repeat_region   26081. .26196
                     /note="MLT1A element fragment"
     repeat_region   26265. .26661
                     /note="LTR2 element fragment"
     repeat_region   26797. .26898
                     /note="MSTC element fragment"
     repeat_region   27949. .28241
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   28319. .28608
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   30032. .30159
                     /note="MER5 element fragment"
     repeat_region   31794. .31980
                     /partial
                     /note="Alu repeat: matches 308. .109 of consensus"
     repeat_region   31981. .32073
                     /partial
                     /note="Alu repeat: matches 93. .1 of consensus"
     repeat_region   32232. .32735
                     /note="MER9 element fragment"
     repeat_region   35168. .35654
                     /note="MER4B element fragment"
     repeat_region   40640. .40693
                     /note="27 copies of 2 mer 94 % conserved"
BASE COUNT    11544 a   7879 c   8141 g  13150 t
ORIGIN


  Query Match           35.4%;  Score 487.6;  DB 9;  Length 40714;
  Best Local Similarity   47.0%;  Pred. No. 1.8e-106;
  Matches  507;  Conservative  225;  Mismatches  311;  Indels   36;  Gaps    4;

Qy    334 ytncaycaytgyaargtnaayathtggytnatgmgnaarcarathytngcnaayaargar 393
          :| | : |  |    ||  |  |:   :|   |:|  |  |  |:       |   :||:
Db  24003 CTGCCCTAGAGATCTGTGGAATTTGACCTTGAGAGAGATGATTTAGGGTATCAGGTGGAA 24062

Qy    394 garathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnttywsn 453
          ||:||::: ||:||:||::: ||:||:||:|| || || || :| :| ||:||  |:::
Db  24063 GAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGTGTCCTGTTAAAGGCACTCAGT 24122

Qy    454 ttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnathaar 513
          ||:  ::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||: |:
Db  24123 TTTGCAAGGGAAGCAGAACATAAAAGTTCAGAAAATTTTCAGCCTGACAATGTGATAGAA 24182

Qy    514 aaraaraaycc-nttywsngarggnaarttyaarytngcngcngarathtgyathtgyaa 572
          ||:||:| :||  ||::: || |  ||:|| ||::  || || ||:||:||:||: |:||
Db  24183 AAGAAAACCCCATTTTTCTGATGATAAATTGAAGCCAGCTGCAGAAATTTGCATAAGTAA 24242

Qy    573 ygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycarmg 632
          :||:||::  ||:|| ||:|   |:||: |:|| ||:||:||:::   | | ||:||::|
Db  24243 CGAGGAGCCAAATGTTAATCACTAAGACCATGGGGAAAATATCTCCATGGCTTGTCAGAG 24302

Qy    633 nwsnwsncarcarwsnathaarwsnyt-ngcntggmgnccnmgnmgnaartggttytgyg 691
          :: :: | :||: :  |:| ::: :|   |   |:| || :| :| ||:|||||:| :|
Db  24303 GTCTTCACGGCAGCCCCTCACATCACTGCCCCCAGAGGCCTAGGAGAAAATGGTTTTATG 24362

Qy    692 gnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtnccng 751
          |  | || || || :: :  ||:| :|| ||: | :| ||::| || ||  |:|| || |
Db  24363 GGTCGGGCCCAGGGTCCCCATGCTATGTGCAGTCTAGGGACTTGGTGCCCCGTGTCCCAG 24422

Qy    752 tnaa--------------------------------ywsngcngtngcnwsngarggng 778
            |                                  ::: || || || ::  |:|  |
Db  24423 CCACTCCAGCCATAACTGAAAGGGGCCAAGGTATAGCTTGGGCTGTTGCTTCAAAGGATG 24482

Qy    779 cnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnm 838
          :: || | :|| ||||||:| || :: || || ||:|| |  || ||   :||::: :
Db  24483 GAAGCCCCAGGCCTTGGCAGCTTCCATGTGGTGTTGAGCCTGCAGGTGCACTGAAGTCAA 24542

Qy    839 gnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggatgc 898
          | ||:||:|| |||||:|| ||   ::| ||:||:||:|| ||:|| ||: | |||||||
Db  24543 GAATTGAGGTTTGGGAACCTCCACCTAGATTTCAGAAGATGTATGGAAATGCCTGGATGC 24602

Qy    899 cnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtnc 958
          | :| ||:||:||:|| |  || |   | :: :: |||:| || :: |  :| || ||
Db  24603 CCAGGCAAAAGTTTGCTGCAGGGGCGAGGACCTCATGGAGAACCTCTGTTAGGGTGGTGT 24662

Qy    959 araarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnm 1018
          :||:|| || || || |  '|:|| || ||::| || || ::  | |  ||  : ::
Db  24663 GGAAGGGAAAAGTGGGGTCATAGCCGCCACACAGAGTCCCTACCAGGGGACCACCTAGTG 24722

Qy   1019 gngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsny 1078
          | || || :| :|  : ?| || :: :: :| :  ||:|| || :| :: || ||::: :
Db  24723 GAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAATGGTAGATCCACCGACAGCC 24782

Qy   1079 tncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggna 1138
          | || | :|| || | :||: :  | ||:|| ||:||:||:|| || ||:|| || || |
Db  24783 TGCACCGTGTGCCTGGAAAAGCCGCAGACACTCAATGCCAGCCTGTGAAAGCAGCAGGGA 24842

Qy   1139 tggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtayy 1198
           |||: : || || | :||: |     |  || ||::|  | ||:|   | || | : |::
Db  24843 GGGAGGCTGTACCCTGCAAAGCCACAGGGGCAGAGCTGCCCAAGATCATGGGAACCAATC 24902

Qy   1199 tnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcny 1258
          | :| ||:   ||:||::| ||: | :| ||:|| || ||: |  |:|||:||:|| || :
Db  24903 TCTTGCATCAGCATGACCTGGATATGAGACATGGAGTCAAAGGAAATCATTTTGGAGCTT 24962

Qy   1259 tnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgytty 1317
          | :| ||:||:||:|| || || | ::  || | :|||| || ||  || :| ||:||:
Db  24963 TAAGATTTGACTGTCCCACTGGATCTCAGACTTCCATGGGGCCTGTAGCCTCTTTGTTTT 25022

Qy   1318 ggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytgyat 1376
          || ||:| : |:|| ||:|| |  || || ||:|  ||:||::| ||::  ||:||:||
Db  25023 GGCCAATATCTCCCATTTGGAATGGCTGTATTTATCCAATGCCTGTACCCCCATTGTAT 25081


RESULT  15
AP004219/c
LOCUS       AP004219               159712 bp    DNA     linear   PRI 26-DEC-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 8q23, clone: KB1222D11,
            complete sequence.
ACCESSION   AP004219
VERSION     AP004219.2  GI:18146615
KEYWORDS    .
```

```
SOURCE      Homo sapiens pre-pro-B cell cell_line:FLEB 14 - 14 DNA,
            clone_lib:Keio BAC library clone:KB1222D11.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159712)
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2001) Nobuyoshi Shimizu, Keio University, School
            of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
            160-8582, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp,
            Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT     On Jan 14, 2002 this sequence version replaced gi:15824050.
FEATURES             Location/Qualifiers
     source          1. .159712
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23"
                     /clone="KB1222D11"
                     /cell_line="FLEB 14 - 14"
                     /cell_type="pre-pro-B cell"
                     /clone_lib="Keio BAC library"
     repeat_region   1. .4242
                     /rpt_family="L1PA2"
                     /evidence=not_experimental
     repeat_region   4668. .4733
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   5899. .5921
                     /rpt_family="(TGGA)n"
                     /evidence=not_experimental
     repeat_region   complement(6205. .6406)
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   6410. .6589
                     /rpt_family="FAM"
                     /evidence=not_experimental
     repeat_region   6619. .6648
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(6649. .6776)
                     /rpt_family="AluJo/FRAM"
                     /evidence=not_experimental
     repeat_region   6860. .7158
                     /rpt_family="AluJo"
                     /evidence=not_experimental
     repeat_region   9083. .9127
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   9188. .9226
                     /rpt_family="(TA)n"
                     /evidence=not_experimental
     repeat_region   complement(9227. .9522)
                     /rpt_family="AluSq"
                     /evidence=not_experimental
     repeat_region   9836. .9933
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   10728. .10762
                     /rpt_family="L1MA1"
                     /evidence=not_experimental
     repeat_region   10763. .11062
                     /rpt_family="AluSg"
                     /evidence=not_experimental
     repeat_region   11063. .11300
                     /rpt_family="L1MA1"
                     /evidence=not_experimental
     repeat_region   11301. .11324
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   11520. .11545
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   12507. .12560
                     /rpt_family="(TAAAAA)n"
                     /evidence=not_experimental
     repeat_region   12793. .12825
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   15792. .15852
                     /rpt_family="AluJ/FRAM"
                     /evidence=not_experimental
     repeat_region   15993. .16013
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(16014. .16312)
                     /rpt_family="AluSq"
                     /evidence=not_experimental
     repeat_region   17126. .17436
                     /rpt_family="AluSg1"
                     /evidence=not_experimental
     repeat_region   complement(17798. .17960)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   18007. .18197
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   18198. .18489
                     /rpt_family="AluSc"
                     /evidence=not_experimental
     repeat_region   18490. .18715
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   complement(18801. .18936)
                     /rpt_family="FLAM_C"
                     /evidence=not_experimental
     repeat_region   complement(19119. .19294)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   19346. .19541
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   19687. .19844
                     /rpt_family="L1MEc"
                     /evidence=not_experimental
     repeat_region   complement(19954. .20081)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   20082. .20106
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   20423. .20566
                     /rpt_family="FRAM"
                     /evidence=not_experimental
     repeat_region   20567. .20636
                     /rpt_family="GA-rich"
                     /evidence=not_experimental
     repeat_region   complement(20973. .21282)
                     /rpt_family="AluY"
                     /evidence=not_experimental
     repeat_region   22379. .22724
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   22933. .23019
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   23044. .23270
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   complement(23275. .23586)
                     /rpt_family="AluSg1"
                     /evidence=not_experimental
     repeat_region   complement(23946. .24035)
                     /rpt_family="MER94"
                     /evidence=not_experimental
```

```
     repeat_region   24491. .24777
                     /rpt_family="AluY"
                     /evidence=not_experimental
     repeat_region   24782. .24794
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   24795. .24909
                     /rpt_family="L1PA4"
                     /evidence=not_experimental
     repeat_region   24910. .24939
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(24981. .25276)
                     /rpt_family="AluJb"
                     /evidence=not_experimental
     repeat_region   complement(27879. .28173)
                     /rpt_family="AluSp"
                     /evidence=not_experimental
     repeat_region   complement(28598. .29054)
                     /rpt_family="MER63B"
                     /evidence=not_experimental
     repeat_region   30710. .30738
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(30739. .30993)
                     /rpt_family="AluJb"
                     /evidence=not_experimental
     repeat_region   30994. .31032
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(31408. .31724)
                     /rpt_family="AluJo"
                     /evidence=not_experimental
     repeat_region   complement(31949. .32254)
                     /rpt_family="AluSg"
                     /evidence=not_experimental
     repeat_region   32336. .32362
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(32463. .32565)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   32685. .32898
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   32995. .33020
                     /rpt_family="(TAAA)n"
                     /evidence=not_experimental

  Query Match             35.4%;  Score 487.6;  DB 9;  Length 159712;
  Best Local Similarity   48.0%;  Pred. No. 2.9e-106;
  Matches  484;  Conservative  210;  Mismatches  293;  Indels   22;  Gaps    3;

Qy     390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
           :||:||: |: : ||:||:||::: ||:||:||: | || ||||| :  :| ||:||  |
Db 138232 AGAAGAAGTTCCTAAGCAGCAAAGCATTCAAGAGATAACTTGGGTGTCATTAAAGGCACT 138173

Qy     450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
           ::: ||:||::| ||:|| ||:||:||::: :  ||:||::  || || ||: |:|  ||
Db 138172 CAGTTTTATAAGGGAAGCAGAGCATAAAACTTTGGAAAATTCGCAGCCTGACTATGCAAT 138113

Qy     510 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 569
           : |:||:||:||:|| ||::: | :|  ||:||:||::  || || ||:||:||:| : |
Db 138112 AGAAAAGAAAAACCCATTTTCTGGGGAGAAATTCAAGCCGGCTGCAGAAATTTGCACAAG 138053

Qy     570 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 629
           :| : |:||::  ||:|| ||:|| ||:||: |:|| ||:||: |:::  || | ||:||
Db 138052 TAGCAAGGAGCCAAATGTTAATCCCCAGGACCATGGGGAAAATGTCTCCAGGCCATGTCA 137993

Qy     630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnc---cnmgnmgnaartggtt 686
           ::|  : :  ||:||: :  |:  ::: :  || |||:| |   | :| :  ||:|||||
Db 137992 GAGACCTTTACAGCAGCCCCTCCCATCATAAGCCTGGAGGCCTACGAGGAAAAAGTGGTT 137933

Qy     687 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 746
           :  :||  | || || || :: :|  |:||:|| ||:|| :| ||::| ||  | ||:||
Db 137932 TCATGGGCCTGGCCCAGGGTCCCTGGGCTGTGTGCAGCCTAGGGACTTGGTGACCTGTGT 137873

Qy     747 nccngtnaaywsngcngtngcnwsngargg----------------ngcnwsnccnaa 788
           || |  | : : || || ||      :||                    |   :  |  |
Db 137872 CCCAGCCACTCCAGCCGTGGCTGAAAGGGGCCAGTGTACAGCTTGGGCTGTGGCTTCAGA 137813

Qy     789 rccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargt 848
           : | |||||::| ||  : || || ||:|| || || ||  |:||::: :| ||:||:||
Db 137812 GGCTTGGCAGCTTCCACCTGGTGTTGAGCCTGTGGGTGCACAGAAGTCAAGAATTGAGGT 137753

Qy     849 ntgggarccnccnathmgnttycaraarathtayggnaaycccntggatgccnmgncaraa 908
           |||||:|| |    ::| ||:||:||:|| ||:|| ||: | |||||||| :| ||:||
Db 137752 TTGGGAACCTCTGCCTAGATTTCAGAAGATGTATGGAAATGCCTGGATGCCCAGGCAAAA 137693

Qy     909 rttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaa 968
           :||:|| |  || || ||  : :: |||:| || :: || :| |  ||   :||:|  ||
Db 137692 GTTTGCTGCAGGGGTGGGGCCCTCATGGAGAACCTCTGCTAGGGCAGTGAGGAAGGAAAA 137633

Qy     969 ygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmg 1028
           :|| || |  ||  | || ||::| || || :: || || |   : ::  | || || :|
Db 137632 TGTGGGGTTCGATTCCCCACACAGAGTCCCTACTGGGGCACTGCCTAGTGGAGCTGTGAG 137573

Qy    1029 nmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygt 1088
           :|  : || || :: :  :| :  ||:|| || :| :: || ||::: :| || ||:||
Db 137572 AAGAGGGCCACCATCCTTCAGACCCCAGAATGGTAGATCCACTGACAGCTTGCACCATGT 137513

Qy    1089 nccngaraarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngt 1148
           || | :||:::  | ||:|  ||:||:||:||  | ||:|| ||  | | | |: :  |
Db 137512 GCCTGGAAAAACTGCAGACATTCAATGCCAGCCCATGAAAGCAGCCAGGAGGAAGGCTAT 137453

Qy    1149 nccntayaaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytg 1208
           || | :||: |    |  || ||::|  | ||:||  |  | | :||::| :| ||:
Db 137452 ACCCTGCAAAGCCACAGGGGCAGAGCTGCCCAAGACCATGAGAACCTACCTTTTGCATCA 137393

Qy    1209 ycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttyga 1268
           | :||::| ||:|| :| | :|   | ||: | ||:||:||:|| || :| :  ||:|
Db 137392 GCGTGACCTGGATGTGAGACCTGTAATCAAAGGAGATCATTTTGGAGCTTTAAAATTTGG 137333

Qy    1269 ytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyt 1327
           :||:|| ||  | ||: | || | :||||| || ||  || :| ||:||:||  |:||:|
Db 137332 CTGCCCCACTTGTTTTTGGACTTCCATGGGCCCTGTAACCTCTTTGTTTTGGCTAATTTT 137273

Qy    1328 tyccnttyggnacngcngtnttyacncartgyytntayytncaytgyat 1376
           |:|| ||:|| || |  |  ||: | ||:| ::| ||::  ||:||:||
Db 137272 TCCCATTTGGAACAGTAGCGTTTCCCCAATACCTGTACTCCCATTGTAT 137224


Search completed: July  8, 2002, 18:08:58
Job time: 18594 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 12:53:47 ; Search time 481.64 Seconds
(without alignments)
4908.627 Million cell updates/sec

Title: US-09-997-610-3
Perfect score: 1377
Sequence: 1 athgtngtnathccngtnyt.....gyytntayytncaytgyatg 1377

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	471.4	34.2	2590	22	AAH14327	Human cDNA sequenc
c	2	452	32.8	3093	22	AAS03687	Rhesus gene locus:
c	3	452	32.8	9236	22	AAS03689	Rhesus gene locus:
c	4	452	32.8	9236	22	AAS03690	Rhesus gene locus:
c	5	452	32.8	9241	22	AAS03688	Rhesus gene locus:
	6	429.8	31.2	3203	22	AAD05134	Human secreted pro
	7	427.6	31.1	6063	22	AAL03635	Human reproductive
	8	427.2	31.0	22509	22	AAS44505	Human LEKTI DNA cl
	9	427.2	31.0	31529	22	AAS44506	Human LEKTI DNA cl
	10	426.4	31.0	6063	22	AAL03634	Human reproductive
	11	421.4	30.6	1736	22	AAH13678	Human cDNA sequenc
	12	421	30.6	465237	24	ABA90193	Human oestrogen re
	13	409	29.7	1485	23	AAS91984	DNA encoding novel
	14	408.8	29.7	1278	22	AAD05178	Human secreted pro
c	15	400.4	29.1	1580	23	AAS87262	DNA encoding novel
c	16	395.2	28.7	7726	22	ABA21079	Human nervous syst
c	17	393.6	28.6	32127	22	AAI99255	Human excretory re
c	18	393.6	28.6	32127	22	AAI63605	Human kidney relat
c	19	393.6	28.6	72215	22	AAK86832	Human immune/haema
c	20	388.2	28.2	1405	22	AAK70265	Human immune/haema
c	21	386.6	28.1	1405	22	AAK70266	Human immune/haema
	22	386.4	28.1	8387	22	ABA14892	Human nervous syst
	23	386.4	28.1	8387	22	AAK84548	Human immune/haema
	24	384.4	27.9	1621	23	AAS67924	DNA encoding novel
	25	378.8	27.5	3049	22	AAH16637	Human cDNA sequenc
c	26	373.8	27.1	81369	21	AAA97997	Human T gene DNA.
	27	371.6	27.0	1306	23	AAS64678	DNA encoding novel
c	28	361.6	26.3	22756	22	AAS40321	DNA encoding human
c	29	361.6	26.3	22756	22	AAL03921	Human reproductive
	30	355.4	25.8	1946	22	ABA44227	Human breast cell
	31	355.4	25.8	1946	22	ABA54677	Human foetal liver
	32	355.4	25.8	1946	22	ABA24459	Probe #2925 for ge
	33	355.4	25.8	1946	22	AAK02964	Human brain expres
	34	355.4	25.8	1946	22	AAK28410	Human bone marrow
	35	355.4	25.8	1946	22	AAI12975	Probe #2908 for ge
	36	355.4	25.8	1946	22	AAI34334	Probe #3020 used t
	37	355.4	25.8	1946	22	AAI02893	Probe #2884 used t
c	38	347	25.2	26410	22	AAK70623	Human immune/haema
c	39	346.8	25.2	32986	22	AAK69758	Human immune/haema
c	40	346.8	25.2	32986	22	AAK84629	Human immune/haema
	41	346.4	25.2	923	23	AAS66623	DNA encoding novel
	42	341.4	24.8	50000	24	AAD26400	Human glutamate re
	43	341.4	24.8	50000	24	AAD26437	Human GRM3 gene fr
c	44	337.4	24.5	57728	22	AAC87588	Human 9p11 chromos
	45	336.6	24.4	2197	22	AAH18389	Human cDNA sequenc

ALIGNMENTS

RESULT 1
AAH14327
ID AAH14327 standard; cDNA; 2590 BP.
XX
AC AAH14327;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:11697.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.

```
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11697; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2590 BP; 722 A; 526 C; 670 G; 672 T; 0 other;

  Query Match             34.2%;  Score 471.4;  DB 22;  Length 2590;
  Best Local Similarity   47.0%;  Pred. No. 1.1e-108;
  Matches  481;  Conservative  213;  Mismatches  293;  Indels   37;  Gaps    3;

Qy     390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
           :||:||:||::: ||:||:||::: ||:||:||:|| |  ||||| :| :| | :|| ||
Db     608 agaagaaatttctaagcagcaaagcattcaagaggtgagttgggtactgttaagagcatt 667

Qy     450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
           ::: ||:| ::| ||: | ||:||:||::  :: ||:||::|  | || ||: |:|  ||
Db     668 cagtttcaaaagggaaacagagcataaaaattcagaaaatttgtagcctgactatgcaat 727

Qy     510 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 569
           : |:||:||:||:|  ||::: | :|  ||:||:||::  || || ||:||:||:||: |
Db     728 agaaaagaaaaaccagttttctggggagaaattcaagccagctgcagaaatctgcataag 787

Qy     570 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 629
           :| : |: |::  ||:|| ||:|| ||:||: |:|| ||:||: |:::  || | ||:||
Db     788 catcaagaagcctaatgttaatccccaggaccatggggaaaatgtctccaggccatgtca 847

Qy     630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naartggtt 686
           ::|  : :: ||:||: :  |:  ::: :  ||   |:|  | :| :|    ||:|||||
Db     848 gagaccttcacagcagcccctcctgtcacaggcccagagatccaggaggaaaaaatggtt 907

Qy     687 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 746
           :| :||  | || || || :: :| ||:||:|| ||:|| :  ||::| || || ||:||
Db     908 ttatgggccaggaccagggtccctgtgctgtgtgcagcctatagacttggtgccctgtgt 967

Qy     747 nccngtnaayws----------------------------------ngcngtngcnwsnga 773
            || |    : :                                  || || || :: ||
Db     968 cccagctgctccagccatggctgaaaggggccaatgtagagctcatgctgtggcttcaga 1027

Qy     774 rggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraa 833
           :|| |  :: || ||:|| |||||::| ||    || || ||:|| || ||  |  |:||
Db    1028 gggtggaagccccaagccttggcagcttccacatggtgttgagcctgtaggtacacagaa 1087

Qy     834 rwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntg 893
           ::: :| ||:||:|| |||||:|| ||   ::| ||:||:||:|| ||:|| ||: | ||
Db    1088 gtcaagaattgaggtttgggaacctccacctagatttcagaagatgtatggaaatgcctg 1147

Qy     894 gatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngt 953
           |||| | :| ||:||:||:|| |  |  |  ||  : :: |||:| || :: || :| |
Db    1148 gatgtccaggcaaaagtttgctgcagtggcagggccctcatggagaacctctgctagggc 1207

Qy     954 ngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnws 1013
             | ||:||:|| | :|| ||   |||:|  || | ::| || || :  || ||    ::
Db    1208 aatgcagaagggaattgtggggccggagctcccacgtagagtccctaatggggcattatc 1267

Qy    1014 nwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacnga 1073
            ::  | || || :| :|  : || || :: :  :| :  ||: | ||  | :: || ||
Db    1268 tagtggagctgtgagaagagggccaccgtccttcagaccccagcatggtggatccactga 1327

Qy    1074 ywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngc 1133
           ::: :| || ||:|  || | :||: :  | ||:|| ||:||:||: | || ||:|| ||
Db    1328 cagcttgcaccatgcgcctggaaaagccgcagacactcaatgccagtccgtgaaagcagc 1387

Qy    1134 nggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnat 1193
             | | |||: : || || | :||: |    |  || ||::|  | ||: |  | ||
Db    1388 caggagggaggctgtaccctgcaaggccacaggggcagaactgccaaaggccatgggtcc 1447

Qy    1194 htayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttygg 1253
           : |::| :| ||:   | :||::|  |:|| :| | : | || ||: | ||:||:||:||
Db    1448 ccacctttgcatcagcgtgacctgaatgtgagacctagagtcaaaggagatcattttgg 1507

Qy    1254 ngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccng-tnccnytnt 1312
             || :| :  ||:||:||:||  | || ||: | || | :|||| || |   || :| |
Db    1508 agctttaaaatttgattgtcctgctggattttggacttgcatgggtcctgcagccccttt 1567

Qy    1313 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1372
           |:||:|| ||:| : |:|| ||:|| |  || || ||:|| ||:| ::| ||::   |:
Db    1568 gttttggccaatgtctcccatttggaatggctgtatttacccaatacctgtacccctatc 1627

Qy    1373 gyat 1376
           |:||
Db    1628 gtat 1631


RESULT   2
AAS03687/c
ID   AAS03687 standard; DNA; 3093 BP.
XX
AC   AAS03687;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Rhesus gene locus: RHD gene deletion in Rh negative haplotypes.
XX
KW   Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW   RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW   haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   primer_bind      32..54
FT                    /*tag=  a
FT                    /note= "Binding site of primer rez7"
FT   primer_bind      complement (3034..3054)
FT                    /*tag=  b
FT                    /note= "Binding site of primer rnb31"
XX
PN   WO200132702-A2.
XX
PD   10-MAY-2001.
XX
PF   31-OCT-2000; 2000WO-EP10745.
XX
PR   02-NOV-1999;   99EP-0121686.
PR   31-MAY-2000; 2000EP-0111696.
XX
PA   (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
```

```
PI   Flegel WA,  Wagner FF;
XX
DR   WPI; 2001-291052/30.
XX
PT   New nucleic acid molecular structure, useful for detection of common
PT   RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT   and RHCE genes -
XX
PS   Example 10; Fig 5; 135pp; English.
XX
CC   The sequence represents the coding sequence of Rhesus gene locus:
CC   RHD gene deletion in Rh negative haplotypes. The Rhesus genes
CC   locus comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC   Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC   and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC   chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC   RHD deletion in the common RHD negative haplotypes. The sequence has
CC   been used to design primers which are useful for: (1) the specific
CC   detection of the common RHD positive haplotypes in D-negative
CC   individuals; (2) blood group typing; (3) determining whether a patient
CC   can be transfused with RhD negative blood and whether blood is suitable
CC   for transfusion to patients who should not be exposed to antigen C; (4)
CC   assessing the risk of a RhD negative mother of conceiving or carrying an
CC   RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC   women who are Rhesus D negative, where the foetus is not homozygous for
CC   the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ   Sequence 3093 BP; 891 A; 754 C; 619 G; 829 T; 0 other;


  Query Match             32.8%;  Score 452;  DB 22;  Length 3093;
  Best Local Similarity   47.4%;  Pred. No. 1.1e-103;
  Matches  485;  Conservative  212;  Mismatches  282;  Indels   45;  Gaps    5;

Qy     390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
           :| :||:||::: ||:||:||::: ||:||:||:|| || ||||| :| :| ||: | ||
Db    1338 AGGAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGACATT 1279

Qy     450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
           ::: ||:||::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||
Db    1278 CAGTTTTATAAGGGAAGCAGAGCATAAAAGTTCGGAAAATTTGCAGCCTGACAATGTGAT 1219

Qy     510 haaraaraaraa--yccnttywsngarggnaarttyaarytngcngcngarathtgyath 567
           : |:||:||:||  :|| ||::: ||:|  ||:||:||::| || || ||:||:||:|| 
Db    1218 AGAAAAGAAAAATTCCCATTTTCTGAGGAGAAATTCAAGCTGGCTGCAGAAATTTGCATG 1159

Qy     568 tgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgy 627
            |:||: |:||::  ||:|  ||: | ||:||:||:|| ||:||: |:::  || | ||: 
Db    1158 AGTAAC-AGGAGCCAAATGCTAATTCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGT 1100

Qy     628 carmgnwsnwsncarcarwsnathaarwsnytng-cntggmgnccnmgnmgnaartggtt 686
           ||::| :: :   :||: :  |:  ::: :  | |  |  |    :| :  ||:|||||
Db    1099 CAGAGGTCTTTATGGCAACCCCTCCCATCACAGGTCCAGAGGTATCAGGAAAAAATGGTT 1040

Qy     687 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 746
           :||: |  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db    1039 TTGTTGGCCAGGCCCGGGGTCCTCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGCAT 980

Qy     747 nccngtna----------------------------------aywsngcngtngcnwsng 772
            || |  |                                  ::: || || || :: |
Db    979 CCCAGCCACTCCCAACCATGACTGACGGGAGGCAAGGTAGAGCTTGGGCTGTAGCTTCGG 920

Qy     773 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 832
            : | || :: || ||:|| || ||::| || :: || || ||: | |   | ||  |:|
Db    919 GGAGTGCAAGCCCCAAGCCTTGACAGCTTCCATGTGGTGTTGAGACTGCGAGTGCACAGA 860

Qy     833 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 892
           |::: :| ||:| :|| ||| |:||  |   ::| ||  |:| :|| | :|| ||: | |
Db    859 AGTCAAGAATTGGGGTTTGGAAACCTTCGCCTAGATTAAAGAGGATGTGCGGAAATGCCT 800

Qy     893 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 952
           ||||||| :| ||:||:||:|| |  || |  ||  : :: |||:|  | :: || :| |
Db    799 GGATGCCCAGTCAGAAGTTTGCTGCAGGAGCAGGGCCCTCATGGAGATCCTCTGCCAGGG 740

Qy     953 tngtncaraarggnaaygtnggntgggarccncccncaymgngtnccnwsnggngcnccnw 1012
             || ||:||:|| ||:|| || |  ||: | || ||::  || || :: || || ||
Db    739 CAGTGCAGAAGGGAAATGTGGGGTCAGAGACCCCACACACAGTCCCTACTGGGGCACCAC 680

Qy    1013 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1072
           : ::  | || || :| :|  : || ||       :| :  ||:|| || :| :: || |
Db    679 CTAGTGGAGCTGTGAGAAGAGGTCCTCC-------AGACCCCAGAATGGTAGATCCACCG 627

Qy    1073 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1132
           |::: :| || | :|| || | :||: :  | ||:|| ||:||:||:||  | ||:|| |
Db    626 ACAGCTTGCACCGTGTACCTGGAAAAGCTGCAGACACTCAATGCCAGCCCATGAAAGCAG 567

Qy    1133 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1192
           | |  | |||: : || || | :||:      |  || ||::|  | ||:||  | || |
Db    566 CTGAGAGGGAGGCTGTACCCTGCAAAGGTACAGGGGCAGAGCTGCCCAAGACCATGGGAA 507

Qy    1193 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycaytttyg 1252
            : |::  :| ||:   | :||::| ||:|| :|   :|| || ||: | ||:||:||:| 
Db    506 CCCACCCCTTCCATCTGCGTGACCTGGATGTGAGATGTGGAGTCAAAGGAGATCATTTTG 447

Qy    1253 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytnt 1312
           | || :| :| ||:||:||:|| || || ||: | ||   :||||| |     || :| |
Db    446 GAGCTTTAAGATTTGACTGCCCCACTGGATTTTGGACTCTCATGGGCCTGTAGCCTCTTT 387

Qy    1313 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1372
           |:||:|  ||:||: |:|| ||:|| |  || || ||:|| ||:|||::| ||::  ||:| 
Db    386 GTTTTGACCAATTTATCCCATTTGGAATGGCTGTATTTACCCAATGCCTGTACCCCCATT 327

Qy    1373 gyat 1376
           |:||
Db    326 GTAT 323


RESULT   3
AAS03689/c
ID   AAS03689 standard; DNA; 9236 BP.
XX
AC   AAS03689;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Rhesus gene locus: upstream Rhesus box of D-positives.
XX
KW   Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW   RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW   haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS   Homo sapiens.
XX
PN   WO200132702-A2.
XX
PD   10-MAY-2001.
XX
PF   31-OCT-2000; 2000WO-EP10745.
XX
PR   02-NOV-1999;   99EP-0121686.
PR   31-MAY-2000; 2000EP-0111696.
XX
PA   (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PI   Flegel WA,  Wagner FF;
XX
DR   WPI; 2001-291052/30.
XX
PT   New nucleic acid molecular structure, useful for detection of common
PT   RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT   and RHCE genes -
XX
PS   Disclosure; Fig 9; 135pp; English.
XX
CC   The sequence represents the coding sequence of Rhesus gene locus:
CC   upstream Rhesus box of D positives. The Rhesus genes locus
```

```
CC   comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC   Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC   and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC   chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC   RHD deletion in the common RHD negative haplotypes. The sequence has
CC   been used to design primers which are useful for: (1) the specific
CC   detection of the common RHD positive haplotypes in D-negative
CC   individuals; (2) blood group typing; (3) determining whether a patient
CC   can be transfused with RhD negative blood and whether blood is suitable
CC   for transfusion to patients who should not be exposed to antigen C; (4)
CC   assessing the risk of a RhD negative mother of conceiving or carrying an
CC   RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC   women who are Rhesus D negative, where the foetus is not homozygous for
CC   the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ   Sequence 9236 BP; 2467 A; 2319 C; 2000 G; 2450 T; 0 other;


  Query Match             32.8%;  Score 452;  DB 22;  Length 9236;
  Best Local Similarity   47.4%;  Pred. No. 2.7e-103;
  Matches  485;  Conservative  212;  Mismatches  282;  Indels   45;  Gaps    5;

Qy     390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
           :| :||:||::: ||:||:||::: ||:||:||:|| || ||||| :| :| ||: | ||
Db    6014 AGGAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGACATT 5955

Qy     450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
           ::: ||:||::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||
Db    5954 CAGTTTTATAAGGGAAGCAGAGCATAAAAGTTCGGAAAATTTGCAGCCTGACAATGTGAT 5895

Qy     510 haaraaraaraa--yccnttywsngarggnaarttyaarytngcngcngarathtgyath 567
           : |:||:||:||  :|| ||::: ||:|  ||:||:||::| || || ||:||:||:||
Db    5894 AGAAAAGAAAAATTCCCATTTTCTGAGGAGAAATTCAAGCTGGCTGCAGAAATTTGCATG 5835

Qy     568 tgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgy 627
            |:||: |:||::  ||:|  ||: | ||:||:||:|| ||:||: |:::  || | ||:
Db    5834 AGTAAC-AGGAGCCAAATGCTAATTCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGT 5776

Qy     628 carmgnwsnwsncarcarwsnathaarwsnytng-cntggmgnccnmgnmgnaartggtt 686
           ||::| :: :    :||: :  |:  ::: :  | |  |  |    :| :  ||:|||||
Db    5775 CAGAGGTCTTTATGGCAACCCCTCCCATCACAGGTCCAGAGGTATCAGGAAAAAATGGTT 5716

Qy     687 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 746
           :||: |  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db    5715 TTGTTGGCCAGGCCCGGGGTCCTCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGCAT 5656

Qy     747 nccngtna----------------------------------aywsngcngtngcnwsng 772
            || |  |                                   ::: || || || :: |
Db    5655 CCCAGCCACTCCCAACCATGACTGACGGGAGGCAAGGTAGAGCTTGGGCTGTAGCTTCGG 5596

Qy     773 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 832
           : | || :: || ||:|| || ||::| || :: || || ||: | |   | ||  |:|
Db    5595 GGAGTGCAAGCCCCAAGCCTTGACAGCTTCCATGTGGTGTTGAGACTGCGAGTGCACAGA 5536

Qy     833 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 892
           |::: :| ||:| :|| ||| |:|| |   ::| ||  |:| :|| | :|| ||: | |
Db    5535 AGTCAAGAATTGGGGTTTGGAAACCTTCGCCTAGATTAAAGAGGATGTGCGGAAATGCCT 5476

Qy     893 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 952
           ||||||| :| ||:||:||:|| |  || |  ||  : :: |||:|  | :: || :| |
Db    5475 GGATGCCCAGTCAGAAGTTTGCTGCAGGAGCAGGGCCCTCATGGAGATCCTCTGCCAGGG 5416

Qy     953 tngtncaraarggnaaygtnggntgggarccncccncaymgngtnccnwsnggngcnccnw 1012
             || ||:||:|| ||:|| || |  ||: | || ||::  || || :: || || ||
Db    5415 CAGTGCAGAAGGGAAATGTGGGGTCAGAGACCCCACACACAGTCCCTACTGGGGCACCAC 5356

Qy    1013 snwsnmgngcngtnmgnmgnwsnccncccnwsnwsnmgnytncaraarggnmgnwsnacng 1072
           : ::  | || || :| :|  : || ||       :| :  ||:|| || :| :: || |
Db    5355 CTAGTGGAGCTGTGAGAAGAGGTCCTCC-------AGACCCCAGAATGGTAGATCCACCG 5303

Qy    1073 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1132
           |::: :| || | :|| || | :||: :  | ||:|| ||:||:||:||  | ||:|| |
Db    5302 ACAGCTTGCACCGTGTACCTGGAAAAGCTGCAGACACTCAATGCCAGCCCATGAAAGCAG 5243

Qy    1133 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1192
           | |  | |||: : || || | :||:      |  || ||::|  | ||:||  | || |
Db    5242 CTGAGAGGGAGGCTGTACCCTGCAAAGGTACAGGGGCAGAGCTGCCCAAGACCATGGGAA 5183

Qy    1193 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1252
           : |::  :| ||:   | :||::| ||:|| :|   :|| || ||: | ||:||:||:|
Db    5182 CCCACCCCTTCCATCTGCGTGACCTGGATGTGAGATGTGGAGTCAAAGGAGATCATTTTG 5123

Qy    1253 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytnt 1312
           | || :| :| ||:||:||:|| || || ||: | ||   :|||| |     || :| |
Db    5122 GAGCTTTAAGATTTGACTGCCCCACTGGATTTTGGACTCTCATGGGCCTGTAGCCTCTTT 5063

Qy    1313 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1372
           |:||:|  ||:||: |:|| ||:|| |  || || ||:|| ||:||::| ||::  ||:|
Db    5062 GTTTTGACCAATTTATCCCATTTGGAATGGCTGTATTTACCCAATGCCTGTACCCCCATT 5003

Qy    1373 gyat 1376
           |:||
Db    5002 GTAT 4999


RESULT   4
AAS03690/c
ID   AAS03690 standard; DNA; 9236 BP.
XX
AC   AAS03690;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Rhesus gene locus: downstream Rhesus box of D-positives.
XX
KW   Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW   RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW   haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS   Homo sapiens.
XX
PN   WO200132702-A2.
XX
PD   10-MAY-2001.
XX
PF   31-OCT-2000; 2000WO-EP10745.
XX
PR   02-NOV-1999;   99EP-0121686.
PR   31-MAY-2000; 2000EP-0111696.
XX
PA   (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PI   Flegel WA,  Wagner FF;
XX
DR   WPI; 2001-291052/30.
XX
PT   New nucleic acid molecular structure, useful for detection of common
PT   RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT   and RHCE genes -
XX
PS   Disclosure; Fig 10; 135pp; English.
XX
CC   The sequence represents the coding sequence of Rhesus gene locus:
CC   downstream Rhesus box of D positives. The Rhesus genes locus
CC   comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC   Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC   and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC   chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC   RHD deletion in the common RHD negative haplotypes. The sequence has
CC   been used to design primers which are useful for: (1) the specific
CC   detection of the common RHD positive haplotypes in D-negative
CC   individuals; (2) blood group typing; (3) determining whether a patient
CC   can be transfused with RhD negative blood and whether blood is suitable
CC   for transfusion to patients who should not be exposed to antigen C; (4)
CC   assessing the risk of a RhD negative mother of conceiving or carrying an
CC   RhD positive foetus. Anti-D antibodies are useful for treating pregnant
```

```
CC   women who are Rhesus D negative, where the foetus is not homozygous for
CC   the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ   Sequence 9236 BP; 2467 A; 2319 C; 2000 G; 2450 T; 0 other;


  Query Match             32.8%;  Score 452;  DB 22;  Length 9236;
  Best Local Similarity   47.4%;  Pred. No. 2.7e-103;
  Matches  485;  Conservative  212;  Mismatches  282;  Indels   45;  Gaps    5;

Qy     390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
           :| :||:||::: ||:||:||::: ||:||:||:|| || ||||| :| :| ||: | ||
Db    6014 AGGAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGACATT 5955

Qy     450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
           ::: ||:||::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||
Db    5954 CAGTTTTATAAGGGAAGCAGAGCATAAAAGTTCGGAAAATTTGCAGCCTGACAATGTGAT 5895

Qy     510 haaraaraaraa--yccnttywsngarggnaarttyaarytngcngcngarathtgyath 567
           : |:||:||:|| :|| ||::: ||:|  ||:||:||::| || || ||:||:||:||
Db    5894 AGAAAAGAAAAATTCCCATTTTCTGAGGAGAAATTCAAGCTGGCTGCAGAAATTTGCATG 5835

Qy     568 tgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgy 627
           |:||: |:||::  ||:|  ||: | ||:||:||:|| ||:||: |:::  || | ||:
Db    5834 AGTAAC-AGGAGCCAAATGCTAATTCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGT 5776

Qy     628 carmgnwsnwsncarcarwsnathaarwsnytng-cntggmgnccnmgnmgnaartggtt 686
           ||::| :: :    :||: :  |:  ::: :  | |  |  |    :| :  ||:|||||
Db    5775 CAGAGGTCTTTATGGCAACCCCTCCCATCACAGGTCCAGAGGTATCAGGAAAAAATGGTT 5716

Qy     687 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 746
           :||: |  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db    5715 TTGTTGGCCAGGCCCGGGGTCCTCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGCAT 5656

Qy     747 nccngtna----------------------------------aywsngcngtngcnwsng 772
           || |  |                                   ::: || || || :: |
Db    5655 CCCAGCCACTCCCAACCATGACTGACGGGAGGCAAGGTAGAGCTTGGGCTGTAGCTTCGG 5596

Qy     773 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 832
           : | || :: || ||:|| || ||::| || :: || || ||: | |   | ||  |:|
Db    5595 GGAGTGCAAGCCCCAAGCCTTGACAGCTTCCATGTGGTGTTGAGACTGCGAGTGCACAGA 5536

Qy     833 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 892
           |::: :| ||:| :|| ||:||   |   ::| ||  |:| :|| | :|| ||: | |
Db    5535 AGTCAAGAATTGGGGTTTGGAAACCTTCGCCTAGATTAAAGAGGATGTGCGGAAATGCCT 5476

Qy     893 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 952
           ||||||| :| ||:||:||:|| |  || |  ||  : :: |||:|  | :: || :| |
Db    5475 GGATGCCCAGTCAGAAGTTTGCTGCAGGAGCAGGGCCCTCATGGAGATCCTCTGCCAGGG 5416

Qy     953 tngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnw 1012
             || ||:||:|| ||:|| || |  ||: | || ||::  || || :: || || ||
Db    5415 CAGTGCAGAAGGGAAATGTGGGGTCAGAGACCCCACACACAGTCCCTACTGGGGCACCAC 5356

Qy    1013 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1072
           : ::  | || || :| :|  : || ||       :| :  ||:|| || :| :: || |
Db    5355 CTAGTGGAGCTGTGAGAAGAGGTCCTCC-------AGACCCCAGAATGGTAGATCCACCG 5303

Qy    1073 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1132
           |::: :| || | :|| || | :||: :  | ||:|| ||:||:||:||  | ||:|| |
Db    5302 ACAGCTTGCACCGTGTACCTGGAAAAGCTGCAGACACTCAATGCCAGCCCATGAAAGCAG 5243

Qy    1133 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1192
           | |  | |||: : || || | :||:      |  || ||::|  | ||:||  | || |
Db    5242 CTGAGAGGGAGGCTGTACCCTGCAAAGGTACAGGGGCAGAGCTGCCCAAGACCATGGGAA 5183

Qy    1193 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1252
           : |::  :| ||:  | :||::| ||:|| :|   :|| || ||: | ||:||:||:|
Db    5182 CCCACCCCTTCCATCTGCGTGACCTGGATGTGAGATGTGGAGTCAAAGGAGATCATTTTG 5123

Qy    1253 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytnt 1312
           | || :| :| ||:||:||:|| || || ||: | ||   :|||| |     || :| |
Db    5122 GAGCTTTAAGATTTGACTGCCCCACTGGATTTTGGACTCTCATGGGCCTGTAGCCTCTTT 5063

Qy    1313 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1372
           |:||:|  ||:||: |:|| ||:|| |  || || ||:|| ||:||::| ||::  ||:|
Db    5062 GTTTTGACCAATTTATCCCATTTGGAATGGCTGTATTTACCCAATGCCTGTACCCCCATT 5003

Qy    1373 gyat 1376
           |:||
Db    5002 GTAT 4999


RESULT   5
AAS03688/c
ID   AAS03688 standard; DNA; 9241 BP.
XX
AC   AAS03688;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Rhesus gene locus: hybrid Rhesus box of RHD negatives.
XX
KW   Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW   RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW   haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS   Homo sapiens.
XX
PN   WO200132702-A2.
XX
PD   10-MAY-2001.
XX
PF   31-OCT-2000; 2000WO-EP10745.
XX
PR   02-NOV-1999;   99EP-0121686.
PR   31-MAY-2000; 2000EP-0111696.
XX
PA   (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PI   Flegel WA,  Wagner FF;
XX
DR   WPI; 2001-291052/30.
XX
PT   New nucleic acid molecular structure, useful for detection of common
PT   RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT   and RHCE genes -
XX
PS   Disclosure; Fig 8; 135pp; English.
XX
CC   The sequence represents the coding sequence of Rhesus gene locus:
CC   hybrid Rhesus box of RHD negatives. The Rhesus genes locus
CC   comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC   Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC   and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC   chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC   RHD deletion in the common RHD negative haplotypes. The sequence has
CC   been used to design primers which are useful for: (1) the specific
CC   detection of the common RHD positive haplotypes in D-negative
CC   individuals; (2) blood group typing; (3) determining whether a patient
CC   can be transfused with RhD negative blood and whether blood is suitable
CC   for transfusion to patients who should not be exposed to antigen C; (4)
CC   assessing the risk of a RhD negative mother of conceiving or carrying an
CC   RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC   women who are Rhesus D negative, where the foetus is not homozygous for
CC   the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ   Sequence 9241 BP; 2454 A; 2320 C; 2014 G; 2453 T; 0 other;


  Query Match             32.8%;  Score 452;  DB 22;  Length 9241;
  Best Local Similarity   47.4%;  Pred. No. 2.7e-103;
  Matches  485;  Conservative  212;  Mismatches  282;  Indels   45;  Gaps    5;

Qy     390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
           :| :||:||::: ||:||:||::: ||:||:||:|| || ||||| :| :| ||: | ||
```

```
Db   6014 AGGAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGACATT 5955

Qy    450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
          ::: ||:|||::| ||:|| ||:||:||::: :: ||:|||::| || || ||:||:|| ||
Db   5954 CAGTTTTATAAGGGAAGCAGAGCATAAAAGTTCGGAAAATTTGCAGCCTGACAATGTGAT 5895

Qy    510 haaraaraaraa--yccnttywsngarggnaarttyaarytngcngcngarathtgyath 567
          : |:||:||:||  :|| ||::: ||:|  ||:||:||::| || || ||:||:||:|| 
Db   5894 AGAAAAGAAAAATTCCCATTTTCTGAGGAGAAATTCAAGCTGGCTGCAGAAATTTGCATG 5835

Qy    568 tgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgy 627
           |:||: |:||::  ||:|  ||: | ||:||:||:|| ||:||: |:::  || | ||: 
Db   5834 AGTAAC-AGGAGCCAAATGCTAATTCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGT 5776

Qy    628 carmgnwsnwsncarcarwsnathaarwsnytng-cntggmgnccnmgnmgnaartggtt 686
          ||::| :: :    :||: :  |:  ::: :  | |  |  |    :| :  ||:|||||
Db   5775 CAGAGGTCTTTATGGCAACCCCTCCCATCACAGGTCCAGAGGTATCAGGAAAAAATGGTT 5716

Qy    687 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 746
          :||: |  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db   5715 TTGTTGGCCAGGCCCGGGGTCCTCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGCAT 5656

Qy    747 nccngtna----------------------------------aywsngcngtngcnwsng 772
           || |  |                                   ::: || || || :: |
Db   5655 CCCAGCCACTCCCAACCATGACTGACGGGAGGCAAGGTAGAGCTTGGGCTGTAGCTTCGG 5596

Qy    773 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 832
          : | || :: || ||:|| || ||::| || :: || || || ||: | |  | ||  |:|
Db   5595 GGAGTGCAAGCCCCAAGCCTTGACAGCTTCCATGTGGTGTTGAGACTGCGAGTGCACAGA 5536

Qy    833 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 892
          |::: :| ||:| :|| ||| |:|| |  ::| ||  |:| :|| | :|| ||: | |
Db   5535 AGTCAAGAATTGGGGTTTGGAAACCTTCGCCTAGATTAAAGAGGATGTGCGGAAATGCCT 5476

Qy    893 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 952
          ||||||| :| ||:||:||:|| |  || |  ||  : :: |||:|  | :: || :| |
Db   5475 GGATGCCCAGTCAGAAGTTTGCTGCAGGAGCAGGGCCCTCATGGAGATCCTCTGCCAGGG 5416

Qy    953 tngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnw 1012
            || ||:||:|| ||:|| || |  ||: | || ||::  || || :: || || ||
Db   5415 CAGTGCAGAAGGGAAATGTGGGGTCAGAGACCCCACACACAGTCCCTACTGGGGCACCAC 5356

Qy   1013 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1072
          : ::  | || || :| :|  : || ||       :| :  ||:|| || :| :: || |
Db   5355 CTAGTGGAGCTGTGAGAAGAGGTCCTCC-------AGACCCCAGAATGGTAGATCCACCG 5303

Qy   1073 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1132
          |::: :| || | :|| || | :||: :  | ||:|| ||:||:||:||  | ||:|| |
Db   5302 ACAGCTTGCACCGTGTACCTGGAAAAGCTGCAGACACTCAATGCCAGCCCATGAAAGCAG 5243

Qy   1133 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1192
          | |  | |||: : || || | :||:      |  || ||::|  | ||:||  | || |
Db   5242 CTGAGAGGGAGGCTGTACCCTGCAAAGGTACAGGGGCAGAGCTGCCCAAGACCATGGGAA 5183

Qy   1193 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1252
           : |::  :| ||:   | :||::| ||:|| :|   :|| || ||: | ||:||:||:| 
Db   5182 CCCACCCCTTCCATCTGCGTGACCTGGATGTGAGATGTGGAGTCAAAGGAGATCATTTTG 5123

Qy   1253 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytnt 1312
          | || :| :| ||:||:||:|| || || ||: | ||   :||||| |     || :| |
Db   5122 GAGCTTTAAGATTTGACTGCCCCACTGGATTTTGGACTCTCATGGGCCTGTAGCCTCTTT 5063

Qy   1313 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1372
          |:||:|  ||:||: |:|| ||:|| |  || || ||:|| ||:||::| ||::  ||:|
Db   5062 GTTTTGACCAATTTATCCCATTTGGAATGGCTGTATTTACCCAATGCCTGTACCCCCATT 5003

Qy   1373 gyat 1376
          |:||
Db   5002 GTAT 4999


RESULT   6
AAD05134
```

```
ID   AAD05134 standard; cDNA; 3203 BP.
XX
AC   AAD05134;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 14 cDNA clone HISBF60, SEQ ID NO:24.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             153..356
FT                   /*tag= a
FT                   /product= "Human secreted protein"
FT   sig_peptide     153..215
FT                   /*tag= b
FT   mat_peptide     216..353
FT                   /*tag= c
FT                   /product= "Human mature secreted protein"
XX
PN   WO200134769-A2.
XX
PD   17-MAY-2001.
XX
PF   01-NOV-2000; 2000WO-US30040.
XX
PR   05-NOV-1999;   99US-0163580.
PR   30-JUN-2000; 2000US-0215130.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Wei P,  Baker KP,  Fiscella M;
XX
DR   WPI; 2001-308781/32.
DR   P-PSDB; AAE01245.
XX
PT   New isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 1; Page 408-409; 519pp; English.
XX
CC   AAD05121-AAD05203 represent cDNAs corresponding to 24 human secreted
CC   protein genes, and AAE01232-AAE01311 represent the proteins they encode.
CC   AAE01312-AAE01340 represent human secreted protein variants or fragments.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   24 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
```

```
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 3203 BP; 868 A; 650 C; 740 G; 943 T; 2 other;


  Query Match             31.2%;  Score 429.8;  DB 22;  Length 3203;
  Best Local Similarity   46.5%;  Pred. No. 4.8e-98;
  Matches  476;  Conservative  211;  Mismatches  297;  Indels   40;  Gaps    6;

Qy     391 gargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntty 450
           ||:||:||::: ||:||:||::: ||:||:||:|| |  ||||| :| :| ||:|| ||:
Db    1639 gaagaaatttctaagcagcaaagcattcaagaggtgatttgggtgctgttaaaagcattt 1698

Qy     451 wsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnath 510
              ||:| ::| ||: |  |:||:||::  :  ||:||::| || || | : |:|    :
Db    1699 cattttaaaagggaaac--agcataaaaatttggaaaatttgcagccagttgatgcagca 1756

Qy     511 aaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtgy 570
            |: |:| :||:|  ||::  ||:|  ||:||:||::| || || ||:||:||:||: |:
Db    1757 gaagagagaaacctgttttttgaggagaaattcaagctggctgcagaaatttgcataagt 1816

Qy     571 aaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycar 630
            |:  :||::| ||:|| ||:|  ||:||:||:|| ||:||: |:::  || | ||:||
Db    1817 tacagggagctgaatgttaatctccaagacaatggggaaaatgtctccagggcatgtcac 1876

Qy     631 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tggtty 687
            :| :: ::   :||: :  |:  ::: :   |   |:   | :| :| ||:   ||||:
Db    1877 aggtcttcatggcagcccctcccatcacagaccctgaagactaggaggaaaaaatggttt 1936

Qy     688 tgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtn 747
           ||:||  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db    1937 tgtgggccaggcccagggtccccatgctgtgtgcagcctagggacttggtgccctgcatc 1996

Qy     748 ccngtnaaywsngcngtngc------------------------------nwsngarg 775
           || |    : : ||  | ||                              :: ||:|
Db    1997 ccagctgctccagccattgctaaaagggccgaggtacagctctgcccatggtttcagagg 2056

Qy     776 gngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarw 835
           | || :: || ||:|| |||||:| || :: || || ||:|| |  || ||   :||::
Db    2057 gtgcaagccccaaaccttggcagcttccatgtggtgttgagcctgcaggtgcatggaagt 2116

Qy     836 snmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaaygccntgga 895
           : :| ||:||:|| |||||:|| ||   ::| ||:||:||:|| ||:|| | : | ||||
Db    2117 caagaattgaggtttgggaacctccatctagatttcagaagatgtatggaatcacctgga 2176

Qy     896 tgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtng 955
           |||  :| ||:||:||:||    || |  |   : :: |||:| || :: || :| |  |
Db    2177 tgctcaggcaaaagtttgctccaggggcagagccttcatggagaacctctgctagggcag 2236

Qy     956 tncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnw 1015
           |   :||:|| ||:|| || | |||:|| || ||::|  | || :: || || ||  : :
Db    2237 tg-tgaagggaaatgtggggttggagcccccacacagaatccctactggggcaccaccta 2295

Qy    1016 snmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayw 1075
           :  | || || :| :| :: || || :: :  :| :  ||:|| || :| :: || ||::
Db    2296 gtggagctgtgcgaagaaggccaccgtccttcagaccccagaatggtagatccaccgaca 2355

Qy    1076 snytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng-cn 1134
           : :|  | | :|  || | :||: : || ||: |  |: |:||:|| || ||:|| | |
Db    2356 gcttgtaccgtgcacctggaaaagccacagaccctaaacgccagcccgtgaaagcagcca 2415

Qy    1135 ggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1194
           ||  ||| :::  | || | : |: |    |  || ||::|  | ||:||    || | :
Db    2416 ggagtggggactataccctgtgaagccacaggggcagagctgcccaagactaagggaacc 2475

Qy    1195 tayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggn 1254
           ||::| :| ||:  :  :||::| ||:|| :| ||:|| || | : | ||:||:||:||
Db    2476 tacctcttgcatcattgtgacctggatgtgagacatggagtcagaggagatcattttgga 2535

Qy    1255 gcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntg 1313
             |  |    ||:||:||:|   | || ||: | || | :|||| |  ||  |  :| ||
Db    2536 acgntataatttgactgcctcgctggattttggacttgcatggggcttgtaacgcctttg 2595

Qy    1314 yttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1373
           :| :|| ||:||: |:|| ||:|| |  || || ||:|| || | ::| ||::  ||:||
Db    2596 ttgtgggcagtttctcccatttggaatggctgtatttacccattacctgtatccccattg 2655

Qy    1374 yatg 1377
           :|||
Db    2656 tatg 2659


RESULT   7
AAL03635
ID   AAL03635 standard; DNA; 6063 BP.
XX
AC   AAL03635;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6323.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
```

PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEO ID NO 6323; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 6063 BP; 1655 A; 1162 C; 1259 G; 1987 T; 0 other;

Query Match 31.1%; Score 427.6; DB 22; Length 6063;
Best Local Similarity 46.4%; Pred. No. 2.9e-97;
Matches 475; Conservative 211; Mismatches 298; Indels 40; Gaps 6;

```
Qy    391 gargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntty 450
          ||:||:||::: ||:||:||::: ||:||:||:|| |  ||||| :| :| ||:|| ||:
Db   3481 gaagaaatttctaagcagcaaagcattcaagaggtgatttgggtgctgttaaaagcattt 3540

Qy    451 wsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnath 510
            ||:| ::| ||: |  |:||:||::  :  ||:||::| || || | : |:|    :
Db   3541 cattttaaaagggaaac--agcataaaaatttggaaaatttgcagccagttgatgcagca 3598

Qy    511 aaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtgy 570
          |: |:| :||:|  ||::  ||:|  ||:||:||::| || || ||:||:||:||: |:
Db   3599 gaagagagaaacctgttttttgaggagaaattcaagctggctgcagaaatttgcataagt 3658

Qy    571 aaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycar 630
          |:  :||::| ||:|| ||:|  ||:||:||:|| ||:||: |:::  || | ||:||
Db   3659 tacagggagctgaatgttaatctccaagacaatggggaaaatgtctccagggcatgtcac 3718

Qy    631 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tggtty 687
          :| :: ::   :||: :  |:  ::: :   |   |:   | :| :| ||:   |||||:
Db   3719 aggtcttcatggcagcccctcccatcacagaccctgaagactaggaggaaaaaatggttt 3778
```

```
Qy     688 tgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtn 747
           ||:||  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db    3779 tgtgggccaggcccagggtccccatgctgtgtgcagcctagggacttggtgccctgcatc 3838

Qy     748 ccngtnaaywsng------------------------------cngtngcnwsngarg 775
           || |    : : |                              | || |  :: ||:|
Db    3839 ccagctgctccagccattgctaaaagggccgaggtacagctctgcccgtggtttcagagg 3898

Qy     776 gngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarw 835
           | || :: || ||:|| |||||::| || :: || || ||:|| |  || ||   :||::
Db    3899 gtgcaagccccaaaccttggcagcttccatgtggtgttgagcctgcaggtgcatggaagt 3958

Qy     836 snmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaaycccntgga 895
           : :| ||:||:|| |||||:|| ||   ::| ||:||:||:|| ||:|| | : | ||||
Db    3959 caagaattgaggtttgggaacctccatctagatttcagaagatgtatggaatcacctgga 4018

Qy     896 tgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtng 955
           |||  :| ||:||:||:||    || |  |   : :: |||:| || :: || :| |  |
Db    4019 tgctcaggcaaaagtttgctccaggggcagagccttcatggagaacctctgctagggcag 4078

Qy     956 tncaraarggnaaygtnggntgggarccncccncaymgngtnccnwsnggngcnccnwsnw 1015
           |   :||:|| ||:|| || | |||:|| || ||::|  | || :: || || ||  : :
Db    4079 tg-tgaagggaaatgtggggttggagccccccacacagaatccctactggggcaccaccta 4137

Qy    1016 snmgngcngtnmgnmgnwsncccnccnwsnwsnmgnytncaraarggnmgnwsnacngayw 1075
           :  | || || :  :| :: || ||.:: :  :| :  ||:|| || :| :: || ||::
Db    4138 gtggagctgtgcaaagaaggccaccgtccttcagaccccagaatggtagatccaccgaca 4197

Qy    1076 snytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng-cn 1134
           : :|  | | :|  || | :||: : || ||: |  |: |:||:|| || ||:|| | |
Db    4198 gcttgtaccgtgcacctggaaaagccacagaccctaaacgccagcccgtgaaagcagcca 4257

Qy    1135 ggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1194
           ||  ||| :::  | || | : |: |    |  || ||::|  | ||:||    || | :
Db    4258 ggagtggggactatacctgtgaagccacaggggcagagctgcccaagactaagggaacc 4317

Qy    1195 tayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggn 1254
           ||::| :| ||:  :  :|||::| ||:|| :| ||:|| || | : | ||:||:||:||
Db    4318 tacctcttgcatcattgtgacctggatgtgagacatggagtcagaggagatcattttgga 4377

Qy    1255 gcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntg 1313
            | :|    ||:||:||:|   | || ||: | || | :||||| |  ||  |  :| ||
Db    4378 acgttataatttgactgcctcgctggattttggacttgcatggggcttgtaacgcctttg 4437

Qy    1314 yttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1373
           :| :|| || ||: |:|| ||:|| |  || || ||:|| || | ::| ||::  ||:||
Db    4438 ttgtgggcactttctcccatttggaatggctgtatttacccattacctgtatccccattg 4497

Qy    1374 yatg 1377
           :|||
Db    4498 tatg 4501


RESULT   8
AAS44505
ID   AAS44505 standard; DNA; 22509 BP.
XX
AC   AAS44505;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human LEKTI DNA clone CIT978SKB_94F21 contig 11, SPINK5 exons 1-4.
XX
KW   Human; SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI; ds;
KW   serine protease inhibitor; atopic disease; Netherton's syndrome; asthma;
KW   eczema; hayfever; antiasthmatic; antiallergic; antiinflammatory;
KW   dermatological; PCR primer; sequencing primer; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200164747-A1.
XX
PD   07-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-GB00897.
XX
PR   02-MAR-2000; 2000GB-0005098.
PR   03-MAR-2000; 2000GB-0005229.
XX
PA   (ISIS-) ISIS INNOVATION LTD.
XX
PI   Hovnanian A,  Chavanas S,  Cookson W,  Moffat M,  Walley A;
XX
DR   WPI; 2001-582149/65.
XX
PT   Determining susceptibility to atopic disease or carrier status of
PT   Netherton's syndrome in humans by identifying variants of or mutations
PT   in SPINK5, a gene encoding lympho-epithelial Kazal-type related
PT   inhibitor -
XX
PS   Disclosure; Page 88-94; 123pp; English.
XX
CC   Sequences AAS44359-AAS44514 represent the SPINK5 gene, contigs and
CC   fragments of a SPINK5 clone, sequencing primers and PCR primers for
CC   SPINK5. SPINK5 encodes lympho-epithelial Kazal-type related inhibitor
CC   (LEKTI), a serine protease inhibitor. Susceptibility or predisposition to
CC   an atopic disease in a human subject can be detected by screening the
CC   genome for one or more polymorphic variants of SPINK5 gene and/or
CC   expression of a variant LEKTI protein in a tissue. Carrier status of a
CC   subject or development of Netherton's syndrome is diagnosed by screening
CC   for the presence of loss-of-function mutations in the SPINK5 gene. An
CC   expression vector comprising a nucleic acid encoding a serine protease
CC   inhibitor or its functional fragment can be used in screening for
CC   compounds with potential pharmacological activity by determining the
CC   serine protease activity of a protein previously identified as a ligand
CC   of the LEKTI protein. The atopic diseases include Netherton's Syndrome,
CC   asthma, eczema and hayfever.
XX
SQ   Sequence 22509 BP; 6725 A; 4291 C; 4284 G; 7209 T; 0 other;


  Query Match             31.0%;  Score 427.2;  DB 22;  Length 22509;
  Best Local Similarity   47.1%;  Pred. No. 1.1e-96;
  Matches  472;  Conservative  207;  Mismatches  303;  Indels   21;  Gaps    7;

Qy     390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
           :||:||:||::: ||:||:||::: ||:||:||: | || ||||| :| :| ||:|| ||
Db   14187 agaagaaatttctaagcagcagagcattcaagagataacttgggtgctcttaaaggcatt 14246

Qy     450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
           ::: ||: |::| ||:|| ||:||:||::: :  ||:||::| || || ||:| :|| ||
Db   14247 cagctttttaagggaagcagaacataaaagtttggaaaatttgcagcctgacagtgtgat 14306

Qy     510 haaraara--araayccnttywsngarggnaarttyaarytngcngcngarathtgyath 567
           :||:||:|  |:||:|| ||::: ||:|  ||:||:||::  |  ||  |:||:||:||:
Db   14307 aaaaaaaaagaaaatccatttttctgaggagaaattcaagccagttgcataaatttgcata 14366

Qy     568 tgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgy 627
            |:| : |:||::   |:|| ||:|| ||:||:||:|| ||:||: |:::   ||  | :
Db   14367 agtagcaaggagcctgatgttaatccccaagacaatgg-gaaaatgtctccaagagatat 14425

Qy     628 carmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tgg 684
           ||::|  : :    :||: :  |:  ::  :  ||   |:|    :| :| ||:   |||
Db   14426 cagagacctttgtggcagcccctcccattacatgcccagaggtttaggaggaaaaaatgg 14485

Qy     685 ttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgy 744
           ||:| : |  | ||    || :: :| ||:||:|| ||:||  | | :   |   |
Db   14486 ttctctaggccaggcttagggtccctctgctgtgtgcagccctgcgttgaagacactcca 14545

Qy     745 gtnccng----------tnaaywsngcngtngcnwsngarggngcnwsnccnaarccntg 794
           ||  | |          |   ::: ||  | || :: ||:|| || :: |  ||:|| ||
Db   14546 gtggctgaaggggccatggcttgggctatggcttccgagggtgcaagcctgaagccttg 14605

Qy     795 gcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntggga 854
            ||::| ||    || || ||:|| ||  | ||  |:||::: :| | :||:|| ||  |
Db   14606 acagcttccacatggtgttgagcctgtgagtgcacagaagtcaagaactgaggtttgaaa 14665
```

```
Qy    855 rccnccnathmgnttycaraarathtayggnaayccntggatgccnmgncaraarttygc 914
          :|| |    ::| | :|   |:||:||:|| ||: | |||||| | :| ||:||:||: |
Db  14666 acctcagcctagatctc---agatatatggaaatgcctggatgtccaggcagaagttttc 14722

Qy    915 ngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtngg 974
          || || ||  | :: :: |||:| || :: || :| |  || ||:||: | ||:|| |
Db  14723 tgtaggggtgaggtcctcatggagaacctctgctagggcagtacagaagagaaatgtgga 14782

Qy    975 ntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnws 1034
          | |||:|| || ||::| || || :: || || ||     :  | || || :| :|
Db  14783 gtaggagcccccacacagagtccctactgg-gcaccacttggtggagctgttagaagaga 14841

Qy   1035 nccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccnga 1094
          ||  | :: :: :|    ||:||  | :| :: || | ::: :| || ||: | || ||
Db  14842 gccttcatcctccagaacccagaatagcagatccacaggtagcttgcaccatatgcctga 14901

Qy   1095 raarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccnta 1154
          :||: : ||  |:|| ||:||:||:|| || ||:|| || || | |||: : || || |
Db  14902 gaaagccacaaacactcagtgccagcctgtgaaggcagctgggagggaggcggtaccctg 14961

Qy   1155 yaaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycayga 1214
          :||: |    |  |  ||::|  | ||:||  | || | : |:::| :| ||:   ||:||
Db  14962 caaagccacagaggtggagttgcccaagaccatgggaacccacctcttgcatcagcatga 15021

Qy   1215 yytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgycc 1274
          ::| ||: | :| ||: |  | ||::   |:||:||:||  | :| :| ||:||:||:|
Db  15022 cctggatatgagacatagattcaaaatatatcattttggaactttaagatttgactgcct 15081

Qy   1275 nacnggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnt 1333
            | |  ||:   || | :||||   | ||  || :| ||:||:|| ||:| : |:|| |
Db  15082 tcctgaattttcgacttgcatggcatctgtagccccttttgttttggccaatatctcccat 15141

Qy   1334 tyggnacngcngtnttyacncartgyytntayytncaytgyat 1376
          |:|| || || || || || ||:|| :| ||::  ||:||:||
Db  15142 ttggaacagctgtattgacccaatggctgtacccccattgtat 15184
```

```
RESULT   9
AAS44506
ID   AAS44506 standard; DNA; 31529 BP.
XX
AC   AAS44506;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human LEKTI DNA clone CIT978SKB_94F21 contig 8, SPINK5 exon 5.
XX
KW   Human; SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI; ds;
KW   serine protease inhibitor; atopic disease; Netherton's syndrome; asthma;
KW   eczema; hayfever; antiasthmatic; antiallergic; antiinflammatory;
KW   dermatological; PCR primer; sequencing primer; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200164747-A1.
XX
PD   07-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-GB00897.
XX
PR   02-MAR-2000; 2000GB-0005098.
PR   03-MAR-2000; 2000GB-0005229.
XX
PA   (ISIS-) ISIS INNOVATION LTD.
XX
PI   Hovnanian A,  Chavanas S,  Cookson W,  Moffat M,  Walley A;
XX
DR   WPI; 2001-582149/65.
XX
PT   Determining susceptibility to atopic disease or carrier status of
PT   Netherton's syndrome in humans by identifying variants of or mutations
PT   in SPINK5, a gene encoding lympho-epithelial Kazal-type related
PT   inhibitor -
XX
PS   Disclosure; Page 95-103; 123pp; English.
XX
CC   Sequences AAS44359-AAS44514 represent the SPINK5 gene, contigs and
CC   fragments of a SPINK5 clone, sequencing primers and PCR primers for
CC   SPINK5. SPINK5 encodes lympho-epithelial Kazal-type related inhibitor
CC   (LEKTI), a serine protease inhibitor. Susceptibility or predisposition to
CC   an atopic disease in a human subject can be detected by screening the
CC   genome for one or more polymorphic variants of SPINK5 gene and/or
CC   expression of a variant LEKTI protein in a tissue. Carrier status of a
CC   subject or development of Netherton's syndrome is diagnosed by screening
CC   for the presence of loss-of-function mutations in the SPINK5 gene. An
CC   expression vector comprising a nucleic acid encoding a serine protease
CC   inhibitor or its functional fragment can be used in screening for
CC   compounds with potential pharmacological activity by determining the
CC   serine protease activity of a protein previously identified as a ligand
CC   of the LEKTI protein. The atopic diseases include Netherton's Syndrome,
CC   asthma, eczema and hayfever.
XX
SQ   Sequence 31529 BP; 9686 A; 5798 C; 5934 G; 10111 T; 0 other;
```

```
  Query Match             31.0%;  Score 427.2;  DB 22;  Length 31529;
  Best Local Similarity   47.1%;  Pred. No. 1.5e-96;
  Matches  472;  Conservative  207;  Mismatches  303;  Indels   21;  Gaps    7;

Qy    390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
          :|||:||:|||:: ||:|||:||::: ||:||:||: | || ||||| :| :| ||:|| ||
Db  14187 agaagaaatttctaagcagcagagcattcaagagataacttgggtgctcttaaaggcatt 14246

Qy    450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
          ::: ||: |::| ||:|| ||:||:||::: :  ||:||::| || || ||:| :|| ||
Db  14247 cagctttttaagggaagcagaacataaaagtttggaaaatttgcagcctgacagtgtgat 14306

Qy    510 haaraara--araayccnttywsngarggnaarttyaarytngcngcngarathtgyath 567
          :||:||:|  |:||:|| ||::: ||:|  ||:||:|||::  |  ||  |:||:||:||:
Db  14307 aaaaaaaaagaaaatccatttctgaggagaaattcaagccagttgcataaatttgcata 14366

Qy    568 tgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgy 627
          |:| : |:||::   |:|| ||:|| ||:||:||:|| ||:||: |:::   ||  | :
Db  14367 agtagcaaggagcctgatgttaatccccaagacaatgg-gaaaatgtctccaagagatat 14425

Qy    628 carmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tgg 684
          ||::|  : :    :||: :  |:  ::  :  ||   |:|    :| :| ||:   |||
Db  14426 cagagacctttgtggcagcccctcccattacatgcccagaggtttaggaggaaaaaatgg 14485

Qy    685 ttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgy 744
          ||:| : |  | ||    || :: :| ||:||:|| ||:||  | | :   |   |
Db  14486 ttctctaggccaggcttagggtccctctgctgtgtgcagccctgcgttgaagacactcca 14545

Qy    745 gtnccng----------tnaaywsngcngtngcnwsngarggngcnwsnccnaarccntg 794
          ||  | |          |   ::: ||  | || :: ||:|| || :: |  ||:|| ||
Db  14546 gtggctgaaggggccatggcttgggctatggcttccgagggtgcaagcctgaagccttg 14605

Qy    795 gcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntggga 854
          ||::| ||    || || ||:|| ||  | ||  |:||::: :| | :||:|| ||  |
Db  14606 acagcttccacatggtgttgagcctgtgagtgcacagaagtcaagaactgaggtttgaaa 14665

Qy    855 rccnccnathmgnttycaraarathtayggnaayccntggatgccnmgncaraarttygc 914
          :|| |    ::| | :|   |:||:||:|| ||: | |||||| | :| ||:||:||: |
Db  14666 acctcagcctagatctc---agatatatggaaatgcctggatgtccaggcagaagttttc 14722

Qy    915 ngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtngg 974
          || || ||  | :: :: |||:| || :: || :| |  || ||:||: | ||:|| |
Db  14723 tgtaggggtgaggtcctcatggagaacctctgctagggcagtacagaagagaaatgtgga 14782

Qy    975 ntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnws 1034
          | |||:|| || ||::| || || :: || || ||     :  | || || :| :|
Db  14783 gtaggagcccccacacagagtccctactgg-gcaccacttggtggagctgttagaagaga 14841

Qy   1035 nccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccnga 1094
          ||  | :: :: :|    ||:||  | :| :: || | ::: :| || ||: | || ||
```

```
Db  14842 gccttcatcctccagaacccagaatagcagatccacaggtagcttgcaccatatgcctga 14901

Qy   1095 raarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccnta 1154
          :||: : ||  |:|| ||:||:||:|| || ||:|| || || | |||: : || || |
Db  14902 gaaagccacaaacactcagtgccagcctgtgaaggcagctgggagggaggcggtaccctg 14961

Qy   1155 yaaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycayga 1214
          :||: |    |  |  ||::|  | ||:||  | || | : |::| :| ||:   ||:||
Db  14962 caaagccacagaggtggagttgcccaagaccatgggaacccacctcttgcatcagcatga 15021

Qy   1215 yytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgycc 1274
          ::| ||: | :| ||: |  | ||::   |:||:||:||  | :| :| ||:||:||:|
Db  15022 cctggatatgagacatagattcaaaatatatcattttggaactttaagatttgactgcct 15081

Qy   1275 nacnggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnt 1333
            | |  ||:   || | :||||   | ||  || :| ||:||:|| ||:| : |:|| |
Db  15082 tcctgaattttcgacttgcatggcatctgtagccccttttgttttggccaatatctcccat 15141

Qy   1334 tyggnacngcngtnttyacncartgyytntayytncaytgyat 1376
          |:|| || || || || || ||:|| :| ||::  ||:||:||
Db  15142 ttggaacagctgtattgacccaatggctgtacccccattgtat 15184


RESULT  10
AAL03634
ID   AAL03634 standard; DNA; 6063 BP.
XX
AC   AAL03634;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6322.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
```

```
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
PT   is used in preventing, treating or ameliorating a medical condition  -
XX
PS   Disclosure; SEQ ID NO 6322; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 6063 BP; 1656 A; 1163 C; 1262 G; 1982 T; 0 other;

  Query Match             31.0%;  Score 426.4;  DB 22;  Length 6063;
  Best Local Similarity   46.4%;  Pred. No. 5.9e-97;
  Matches  475;  Conservative  210;  Mismatches  299;  Indels   40;  Gaps    6;

Qy     391 gargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntty 450
Db    3481 gaagaaatttctaagcagcaaagcattcaagaggtgatttgggtgctgttaaaagcattt 3540

Qy     451 wsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnath 510
Db    3541 cattttaaaagggaaac--agcataaaaatttggaaaatttgcagccagttgatgcagca 3598

Qy     511 aaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtgy 570
Db    3599 gaagagagaaacctgttttttgaggagaaattcaagctggctgcagaaatttgcataagt 3658

Qy     571 aaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycar 630

Db    3659 tacagggagctgaatgttaatctccaagacaatggggaaaatgtctccagggcatgtcac 3718

Qy     631 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tggtty 687
Db    3719 aggtcttcatggcagcccctcccatcacagaccctgaagactaggaggaaaaaatggttt 3778

Qy     688 tgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtn 747
Db    3779 tgtgggccaggcccagggtccccatgctgtgtgcagcctagggacttggtgccctgcatc 3838

Qy     748 ccngtnaaywsng----------cngtngcnwsngarg 775
Db    3839 ccagctgctccagccattgctaaaagggccgaggtacagctctgcccgtggtttcagagg 3898

Qy     776 gngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarw 835
Db    3899 gtgcaagccccaaaccttggcagcttccatgtggtgttgagcctgcaggtgcatggaagt 3958

Qy     836 snmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntgga 895
Db    3959 caagaattgaggtttgggaacctccatctagatttcagaagatgtatggaatcacctgga 4018

Qy     896 tgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtng 955
Db    4019 tgctcaggcaaaagtttgctccaggggcagagccttcatggagaacctctgctagggcag 4078

Qy     956 tncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnw 1015
Db    4079 tg-tgaagggaaatgtggggttggagcccccacacagaatccctactggggcaccaccta 4137

Qy    1016 snmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayw 1075
Db    4138 gtggagctgtgcaaagaaggccaccgtccttcagaccccagaatggtagatccaccgaca 4197

Qy    1076 snytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng-cn 1134
Db    4198 gcttgtaccgtgcacctggaaaagccacagaccctaaacgccagcccgtgaaagcagcca 4257

Qy    1135 ggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1194
Db    4258 ggagtggggactataccctgtgaagccacaggggcagagctgcccaagactaagggaacc 4317

Qy    1195 tayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggn 1254
Db    4318 tacctcttgcatcattgtgacctggatgtgagacatggagtcagaggagatcattttgga 4377

Qy    1255 gcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntg 1313
Db    4378 acgttataattggactgcctcgctggatttggacttgcatggggcttgtaacgcctttg 4437

Qy    1314 yttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1373
Db    4438 ttgtgggcactttctcccatttggaatggctgtatttacccattacctgtatccccattg 4497

Qy    1374 yatg 1377
Db    4498 tatg 4501


RESULT  11
AAH13678
ID   AAH13678 standard; cDNA; 1736 BP.
XX
AC   AAH13678;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:10539.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
```

PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 10539; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1736 BP; 449 A; 372 C; 453 G; 462 T; 0 other;

Query Match 30.6%; Score 421.4; DB 22; Length 1736;
Best Local Similarity 46.2%; Pred. No. 3.9e-96;
Matches 474; Conservative 211; Mismatches 293; Indels 47; Gaps 6;

```
Qy    390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
          :||:||:||::: ||:||:||::: ||:||:||:|| |  ||| | :| :| ||:|| ||
Db    110 agaagaaatttctaagcagcaaagcattcaagaggtgatttggatactgttaaaggcatt 169

Qy    450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
          ::: ||:||::| ||:|| ||:||:||::: :  ||:||::| || |   |: |:|  ||
Db    170 tagttttataagggaagcagagcataaaagtttggaaaatttgcagcattactatgcgat 229

Qy    510 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 569
          : | ||:||:||:|| ||::: | :|  | :||:||::  || || ||:||:||: |: |
Db    230 agacaagaaaaacccatttctggggagagattcaagccagctgcggaaatttgtgtaag 289

Qy    570 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 629
          :| : |:||::  ||:|| | :|| ||:||: |:|| ||: |: |:::  |  | ||:||
Db    290 tagcaaggagcctaatgttagtccccaagaccatggggaagatgtctccagaccatgtca 349

Qy    630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 689
          ::|  : :: | :||: :  |:  ::: :  ||   |:     :| :  ||:|||||:||
Db    350 gagaccttcaccacagcccctcctatcacaggcccaga-----aggaaaaagtggttttg 404

Qy    690 yggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtncc 749
          :||  | || || || :: :  ||:||:|| ||:|| :| ||: | |  || ||:|| ||
Db    405 tgggcctggtccagggtccccttgctgtgtgcaacctagggatgtggcaccctgtgtccc 464

Qy    750 ngtnaayws----------------------------------ngcngtngcnwsnga 773
          |    : :                                   |  || || :: ||
Db    465 agctgctgcttcagctgtggctgaaaggggccaatgtacagctcaggttgtggcctcaga 524

Qy    774 rggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtngngcnaaraa 833
          :|| |  :: || ||:|  |||||::  ||    || || ||:|| |  || ||  |:||
Db    525 gggtggaagccccaagctttggcagcgtccacatggtgttgagcctgcaggtgcacagaa 584

Qy    834 rws-nmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 892
          :::  :| ||:||:|| |||||:|| |    ::| ||:|   |:|| ||:|| ||:   |
Db    585 gtcaaagaattgaggtttgggaacctctacctagatttc---agatgtatggaaatgtct 641

Qy    893 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 952
          |||||| :| ||:||:||:||  | || || ||  : :: | |:| || :: || :| |
Db    642 agatgcccaggcaaaagtttgctat-ggggtggggccctcatagagaacctctgctaggg 700

Qy    953 tngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnw 1012
            ||   :||:|| ||:|| || | |||:|| || ||::| || || :: || || |
Db    701 cagtgtggaagggaaatgtggggttggagcccccacacagagtccctactggggcactgc 760

Qy   1013 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1072
          : :: :| || || :| :|  : ||    :: :: :| :  ||:|| || :|  : || |
Db    761 ctagtagagctgtgagaagagggccattgtcttctagaccccagaatggcagacccactg 820

Qy   1073 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1132
          |::: :| || ||:|| || | :||: :  | ||:|| ||: |:||:|| || ||:||
Db    821 acagcttgcaccatgtgcctggaaaagctgcagacactcaacgccagcccgtgaaagcat 880

Qy   1133 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1192
          | || | |||: : || || | : |: |    |  || ||::|  | ||:||  | || |
Db    881 ccgggagggaggctgtaccctgccaagccacaggggcagagctacccaagaccatgggaa 940

Qy   1193 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1252
          : |::| :| ||:     :||:|| ||: | :| | :||  | ||: | ||:||:||:|
Db    941 ctcacctcttgcatcagtgtgacctggatttgagacctggaatcaaaggagatcattttg 1000

Qy   1253 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytn 1311
          | || :| :  ||:||:||:|| || |  ||: | || | :||||| || ||  || :
Db   1001 gagctttaacatttgactgccccactgaattctggacttgcatgggacccgtaaccccat 1060

Qy   1312 tgyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncay 1371
          ||:||:|| ||:||: |: | ||:|| |  || || ||:|| ||:| ::|  |:   ||:
Db   1061 tgttttggccaatttctctcatttggaatggctgtatttacccaatacctgcacacccat 1120

Qy   1372 tgyat 1376
          ||:||
Db   1121 tgtat 1125
```

RESULT 12
ABA90193
ID ABA90193 standard; DNA; 465237 BP.
XX
AC ABA90193;
XX
DT 11-FEB-2002 (first entry)
XX
DE Human oestrogen receptor alpha gene.
XX
KW Human; oestrogen receptor alpha; ESR-alpha; ER; chromosome 6; Syne-2;
KW synaptic nuclei expressed gene 2; haplotype; cytostatic; osteopathic;
KW cardiant; vasotropic; gene therapy; vaccine; cancer; osteoporosis;
KW cardiovascular disease; oestrogen receptor; ds.
XX
OS Homo sapiens.
XX

```
PN   WO200162969-A2.
XX
PD   30-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05358.
XX
PR   22-FEB-2000; 2000US-183756P.
PR   20-OCT-2000; 2000US-0692414.
PR   24-JAN-2001; 2001US-0768184.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES;
XX
DR   WPI; 2002-041152/05.
DR   P-PSDB; AAG68251.
XX
PT   Novel variant of estrogen receptor alpha polypeptide useful for
PT   determining the biological activity of a protein for high throughput
PT   screening and for raising antibodies that elicit an immune response in
PT   host -
XX
PS   Example; Fig 1 page 1-93; 333pp; English.
XX
CC   The present invention describes an isolated peptide (I) consisting of an
CC   amino acid sequence selected from: (a) the amino acid sequence of a
CC   variant of the oestrogen receptor alpha (ESR-alpha) protein in AAG68251;
CC   or (b) a fragment comprising at least 10 contiguous amino acids of the
CC   protein in AAG68251. (I) has cytostatic, osteopathic, cardiant and
CC   vasotropic activities, and can be used in gene therapy and vaccine
CC   production. (I) is useful for identifying an agent that binds to (I), by
CC   contacting (I) with an agent and assaying the contacted mixture to
CC   determine whether a complex is formed with the agent bound to the
CC   peptide. A polynucleotide (II), encoding (I), is useful in the
CC   development of diagnostics and therapies for diseases and disorders
CC   mediated/modulated by an oestrogen receptor (ER). (II) is also useful in
CC   gene therapy for treating cancer, osteoporosis and cardiovascular
CC   diseases. The human ESR-alpha gene is located on chromosome 6. The
CC   present sequence represents the human ESR-alpha gene, which is given in
CC   the exemplification of the present invention.
XX
SQ   Sequence 465237 BP; 133988 A; 89578 C; 93946 G; 147721 T; 4 other;
```

```
  Query Match             30.6%;  Score 421;  DB 24;  Length 465237;
  Best Local Similarity   46.5%;  Pred. No. 5e-94;
  Matches  477;  Conservative  213;  Mismatches  292;  Indels   43;  Gaps    7;

Qy     392 argarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnttyw 451
           |:||:||::: ||:||:||::  ||:||: |:|| || ||||| :| :| ||:|| ||::
Db 438788 aagaaatttctaagcagcaaaacattcaaaaggtgacttgggtgttgttaaaagcattct 438847

Qy     452 snttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnatha 511
           : ||:| ::| ||: |  |:||:||::: :: ||:||::| || || ||: |:|   |:
Db 438848 gttttaaaagagaaac--agcataaaagttcagaaaatttgcagcctgatgatgcagtag 438905

Qy     512 araaraaraaycc-nttywsngarggnaarttyaarytngcngcngarathtgyathtgy 570
           :||:||:||:||  ||::  ||:|  ||:||:||::| || || ||:||:||:||: |:
Db 438906 gaaagaaaaacccatttttttgaggagaaattcaagctggctgcagaaatttgcataagt 438965

Qy     571 aaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycar 630
           ||: |:||::  ||:|| ||:|| ||:||:||:|| ||:||: |:::  |  | ||:||
Db 438966 aacaaggagccaaatgttaatccccaagacaatggggaaaatgtctccagagcatgtcat 439025

Qy     631 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg----naartggtt 686
           :| :: ::   :||: :  |:  ::: :   |  ||:  || :| :|     ||:  |||
Db 439026 aggtcttcatggcagcccctcccatcacagacccggaagcctaggaggaaaaaaacagtt 439085

Qy     687 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 746
           :||:||  | |  || || :: :  ||:||:|| ||:|| :|  |::| || || ||: |
Db 439086 ttgtgggccagtcccagggtccccatgctgtgtgcagcctaggaacttggtgccctgcat 439145

Qy     747 nccngtnaaywsngcngtngc----------------------nwsngarggngcnws 782
```

```
           | |    : : ||  | ||                          :: ||:|  || ::
Db 439146 ctcagctgctccagctattgctaaaaggggctgaggtaccacggtttcagaggttgcaag 439205

Qy     783 nccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnat 842
           || ||:|| |||||::|  | :: || || ||:|| ||  |  |  |:||::  :| ||
Db 439206 ccccaaaccttggcagctttcatgtggtgttgagcctgtgtgtacacagaagttaagaat 439265

Qy     843 hgargtntgggarccnccnathmgnttycaraarathtayggnaaycentggatgccnmg 902
           :||:|| |||||:|| ||  |::  ||:||:||:||:  :|| ||: | ||||| || :|
Db 439266 tgaggtttgggaacctccacttatatttcagaagatacgtggaaatgcctggatacccag 439325

Qy     903 ncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraa 962
           ||:| :||:|| |  |  || ||  : :: |||:|  | :: || :| |  |   :||
Db 439326 gcaaacatttgctgcagaggtggggccctcatggagggcctctgctagggcaatgaggaa 439385

Qy     963 rggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngc 1022
           :|| ||:|| || | |||:|| || ||::| || || :: || || |   : ::  | ||
Db 439386 gggaaatgtggggttggaaccccacacagagtccccactggggcactgcctagtggagc 439445

Qy    1023 ngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytnca 1082
           || :|  |  : || |  :: :: :| :  ||:||  | :| :: || ||::: :| ||
Db 439446 tgtaagaggaggaccactgtcctccagaccgcagaatagtagatccactgacagcttgca 439505

Qy    1083 rcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggna---- 1138
           ||:|| || | :||: : || ||:|| ||: |:||:|| || ||:|| |   |
Db 439506 ccatgtgcctggaaaagccacagacactcaacgccagcctgtgaaagcagtcagggttgg 439565

Qy    1139 ------tggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1192
                ||| : :  | || ||:||: |    |  || ||::|  | ||:||  | || |
Db 439566 aggtggtggtggctataccctataaagccacaggggcagagctgcccaagactatgggaa 439625

Qy    1193 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1252
           :||::| :| ||:   ||:||::| ||:|| :| ||:   || ||: | ||: |:||:|
Db 439626 cctacctcttgcatcagcatgacctggatgtgagacattcagtcaaaggagat-attttg 439684

Qy    1253 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytn 1311
             || :|    ||:||:||:||    || ||:   || | : |||| || ||  || :|
Db 439685 aagctttagaatttgactgccctggtggattttagacttgtgtgggccctgtaacccctt 439744

Qy    1312 tgyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncay 1371
           ||:||:|| ||:||: |:|| ||:||  | || || ||:|| ||:||::|  |::  ||:
Db 439745 tgttttggccaatttctcccatttggagctgctgtatttacccaatgcctaaacccgcat 439804

Qy    1372 tgyat 1376
           ||:||
Db 439805 tgtat 439809
```

```
RESULT 13
AAS91984
ID   AAS91984 standard; cDNA; 1485 BP.
XX
AC   AAS91984;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27788.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
```

```
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27797.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 27788; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1485 BP; 442 A; 333 C; 373 G; 337 T; 0 other;


  Query Match             29.7%;  Score 409;  DB 23;  Length 1485;
  Best Local Similarity   46.5%;  Pred. No. 4.8e-93;
  Matches  458;  Conservative  199;  Mismatches  287;  Indels   40;  Gaps    6;

Qy    433 gtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnwsngaraayytn 492
          |  :  :|  | || ||::: ||:||::| ||:|| ||:||:||::  :  ||:||::|
Db    103 gcttctttgtatgcattcagttttataagggaagcagagcataaaaatttggaaaatttg 162

Qy    493 cayccngayaaygtnathaaraaraaraayccnttywsngarggnaarttyaarytngcn 552
          || || ||: |:|  ||: |:||:||:||:|| ||::: | :|  ||: |:||::
Db    163 cagcctgactatgcaatagaaaagaaaaacccatttctggggagaaactcaagcctatt 222

Qy    553 gcngarathtgyathtgyaaygargarytnaaygtnaayccncargayaayggngaraay 612
          || ||:||:||: |: |:| : |:||::  ||:|| ||:|| ||:||: |: | | :||:
Db    223 gcagaaatttgcctaagtagcaaggagcctaatgttaatccccaagaccatcggggaaat 282

Qy    613 athwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccn 672
           |::   || | ||:||: | :: ::   :| : :  |:  ::: :  || || :| ||
Db    283 gtctacaggccatgtcagggaacttcatggctgcccctcccatcacaagcctgcaggccc 342

Qy    673 mgnmg---naartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgn 729
          :| :|    ||:|||||:  :||  | || || || :  :| ||: |:|| ||:|  :|
Db    343 aggagtaaaaagtggtttcatgggccaggcccagggttcctgtgccgtgtgcagcacagg 402

Qy    730 gayytngtnccntgygtnccng----------------------------------tnaay 756
          ||::| || || ||:|| || |                                  :
Db    403 gacttggtgccctgtgtcccagctgctccagctatggctgaaaggggccaaggtagaggt 462

Qy    757 wsngcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngar 816
          :: || || || :: ||:|| || :: || ||:|| |||||::| ||  :  |  || ||:
Db    463 tgggctgtggcttcagagggtgcaagacccaagccttggcagcttccacgtgatgttgag 522

Qy    817 ccngtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraar 876
          || || || ||  |:||::: :| | :||:|| |||||:|| |    ::| ||:||:||:
Db    523 cctgtgggtgcacagaagtcaagaactgaggtttgggaacctctgcctagatttcagaag 582

Qy    877 athtayggnaayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntgg 936
          || ||:|| ||: | |||||||| :| ||:||:||:|| |  || || ||  : :: |||
Db    583 atgtatggaaacgcctggatgcccaggcaaaagtttgctgcaggtgtgggaccctcatgg 642

Qy    937 mgnacnwsngcnmgngtngtncaraarggnaaygtnggntgggar-ccnccncaymgngt 995
          :| || :: || :| |  || ||:||:|  ||:|| || |  ||: || || ||::| ||
Db    643 agaacctctgctagggcagtgcagaaggaaaatgtggggtcagagccccccacacagagt 702

Qy    996 nccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytnca 1055
           || :: || || ||  : ::  | || || :| :|  : || |  :: :: :| :  ||
Db    703 ccctactggggcaccacctagtggagctgtgagaaggggccactgtcctccagacccca 762

Qy   1056 raarggnmgnwsna-cngaywsnytncarcaygtnccngaraarwsnacngayacncart 1114
          :|| || :| :: | |  |::: :| ||   :|  || | :||: : || || || ||:|
Db    763 gaatggtagatctacccaacagcttgcactgtgcacctggaaaagccacagaaactcaat 822

Qy   1115 gycarccngtnaargcngcnggnatggarw-sngtnccntayaaracngtngtngcngar 1173
          |:||: | || ||:||  |  | |  | :: : || || | :||: |    |  |  ||:
Db    823 gccagaccgtgaaagcaactaggagagggtgctgtaccctgcaaagccacaggggtggag 882

Qy   1174 ytnacnaaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggn 1233
          :|  | ||: |  | ||   : |: | :| ||:   ||:||::| ||:|| :| ||:||
Db    883 ctgcccaaggccatgggagcccacgtcttgcatcagcatgacctggatgttagacatgga 942

Qy   1234 gtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntay 1293
          || ||: | ||:||:||:|| || :| :  ||:||:||:|| || || ||::  || | :
Db    943 gtcaaaggagatcattttggggctttaaaatttgactgccctactggatttcagacttgc 1002

Qy   1294 atgggnccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyac 1352
           |||| || ||  || :| ||:||:|| ||:| : |:||  |:|| |  || || ||:||
Db   1003 gtgggccctgtaaccactttgttttggccaatatctcccagttggaatggctgtatttac 1062

Qy   1353 ncartgyytntayytncaytgyat 1376
           ||:| ::| ||::| ||:||:||
Db   1063 tcaatacctgtacctccactgcat 1086


RESULT  14
AAD05178
ID   AAD05178 standard; cDNA; 1278 BP.
XX
AC   AAD05178;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 14 cDNA clone HISBF60, SEQ ID NO:68.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             224..268
FT                   /*tag= a
FT                   /product= "Human secreted protein"
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200134769-A2.
XX
PD   17-MAY-2001.
XX
```

PF 01-NOV-2000; 2000WO-US30040.
XX
PR 05-NOV-1999; 99US-0163580.
PR 30-JUN-2000; 2000US-0215130.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Wei P, Baker KP, Fiscella M;
XX
DR WPI; 2001-308781/32.
DR P-PSDB; AAE01289.
XX
PT New isolated nucleic acid molecule encoding a human secreted protein is
PT used in preventing, treating or ameliorating a medical condition -
XX
PS Claim 1; Page 442; 519pp; English.
XX
CC AAD05121-AAD05203 represent cDNAs corresponding to 24 human secreted
CC protein genes, and AAE01232-AAE01311 represent the proteins they encode.
CC AAE01312-AAE01340 represent human secreted protein variants or fragments.
CC The secreted proteins and their genes are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 24 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
SQ Sequence 1278 BP; 344 A; 264 C; 358 G; 299 T; 13 other;

Query Match 29.7%; Score 408.8; DB 22; Length 1278;
Best Local Similarity 45.8%; Pred. No. 4.8e-93;
Matches 458; Conservative 205; Mismatches 298; Indels 40; Gaps 6;

```
Qy    391 gargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntty 450
          ||:||:||::: ||:||:||::: ||:||:||:|| |  ||||| :| :| ||:|| ||:
Db    280 gaagaaatttctaagcagcaaagcattcaagaggtgatttgggtgctgttaaaagcattt 339

Qy    451 wsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnath 510
            ||:| ::| ||: |  |:||:||::  :  ||:||::| || || | : |:|    :
Db    340 cattttaaaagggaaac--agcataaaaatttggaaaatttgcagccagttgatgcagca 397

Qy    511 aaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtgy 570
           |: |:| :||:|  ||::  ||:|  ||:||:||::| || || ||:||:||:||: |:
Db    398 gaagagagaaacctgtttttgaggagaaattcaagctggctgcagaaatttgcataagt 457

Qy    571 aaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycar 630
           |:  :||::| ||:|| ||:|  ||:||:||:|| ||:||: |:::  || | ||:||
Db    458 tacagggagctgaatgttaatctccaagacaatggggaaaatgtctccagggcatgtcac 517

Qy    631 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tggtty 687
          :| :: ::   :||: :  |:  ::: :   |   |:   | :| :| ||:   |||||:
Db    518 aggtcttcatggcagccctcccatcacagaccctgaagactaggaggaaaaaatggttt 577
```

```
Qy    688 tgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtn 747
          ||:||  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db    578 tgtgggccaggcccagggtccccatgctgtgtgcagcctagggacttggtgccctgcatc 637

Qy    748 ccngtnaaywsngcngtngc----------------------------------nwsngarg 775
          || |    : : ||  | ||                                  :: ||:|
Db    638 ccagctgctccagccattgctaaaagggccgaggtacagctctgcccatggtttcagagg 697

Qy    776 gngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarw 835
          | || :: || ||:|| |||||::| || :: || || ||:|| |  || ||   :||::
Db    698 gtgcaagccccaaaccttggcagcttccatgtggtgttgagcctgcaggtgcatggaagt 757

Qy    836 snmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntgga 895
          : :| ||:||:|| |||||:|| ||   ::| ||:||:||:|| ||:|| | : | ||||
Db    758 caagaattgaggtttgggaacctccatctagatttcagaagatgtatggaatcacctgga 817

Qy    896 tgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtng 955
          |||  :| ||:||:||:||    || |  |   : :: |||:| || :: || :| |  |
Db    818 tgctcaggcaaaagtttgctccaggggcagagccttcatggagaacctctgctagggcag 877

Qy    956 tncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnw 1015
          |   :||:|| ||:|| || | |||:|| |: ||::|  | || :: || || ||  : :
Db    878 tg-tgaagggaaatgtggggttggagccccyacacagaatccctactggggcaccaccta 936

Qy   1016 snmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayw 1075
          :  | || || :| :| :: || || :: :  :| :  ||:|| || :| :: || ||::
Db    937 gtggagctgtgcgaagaaggccaccgtccttcagaccccagaatggtrgatccaccgaca 996

Qy   1076 snytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng-cn 1134
          : :|  | | :|  || | :||: : || ||: |  |: |:||:|| || ||:|| | |
Db    997 gcttgtaccgwgcacctggaaaagccacagaccctaaacgccagcccgtgaaagcagcca 1056

Qy   1135 ggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1194
          ||  ||| :::  | || | : |: |    |  || ||::|  | ||:||   || | :
Db   1057 ggagtggggactataccctgtgaagccacaggggcagagctgcccaagactaagggaacc 1116

Qy   1195 tayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggn 1254
          ||::| :| ||:  :   ||::|  |:|| :| ||:|| || | : | ||:||:||:|
Db   1117 tacctcttgcatcattgngacctgnatgtgagacatggagtcagaggagatcattttgna 1176

Qy   1255 gcnytnmgnttygaytgyccnacnggnttymgnacntayatgggncngtnccnytntgy 1314
           |  |    ||:||:||:|   | || ||: | || | :|||| |  |   | :  | :
Db   1177 acgntataatttgactgcctcgctggattttggacttgcatggggcttgnaacgcctttt 1236

Qy   1315 ttyg-gncarttyttyccnttyggnacngcngtnttyacnc 1354
             |:| | ||:||: |:|| ||:|| |  || || ||:|| |
Db   1237 gttgtggcagtttctcccatttggaatggctgtatttaccc 1277
```

RESULT 15
AAS87262/c
ID AAS87262 standard; cDNA; 1580 BP.
XX
AC AAS87262;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #23066.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG23075.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 23066; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1580 BP; 357 A; 412 C; 328 G; 483 T; 0 other;

Query Match 29.1%; Score 400.4; DB 23; Length 1580;
Best Local Similarity 42.4%; Pred. No. 7.6e-91;
Matches 466; Conservative 215; Mismatches 383; Indels 34; Gaps 4;

```
Qy     307 aaytaytaywsnwsnttygaygtngarytncaycaytgyaargtnaayathtggytnatg 366
           ||:| : |::        ||:   |   | |    : |: | || ||:  :  |:
Db    1107 AATTCCGATAACGAACAAGACTGTGGCATGCTAACTAGTTACGTGAACCCCGAGCAGTCA 1048

Qy     367 mgnaarcarathytngcnaayaargargarathwsnaarcarcarwsnathcargargtn 426
               |  | ::|     |   : |:| :  ::  |  || | :::  |     |
Db    1047 GCATCCCCCAACTTCTTAGAGGGACAAGTGGCATTCAGCCACCTGAGATTGAGCAATAAC 988

Qy     427 acntgggtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnwsngar 486
           |  || |  :|      :|   |::  ||:| ::| ||:|| ||:||:||::  :  | :
Db     987 AGATGTGGTTTGGAATTGGAACTTATATTTAAAAGGGAAGCAGAGCATAAAAAAATTGGA 928

Qy     487 aayytncayccngayaaygtnathaaraaraaraayccnttywsngarggnaarttyaar 546
           ||::| || |  ||:| :|  ||: |:||:||:||:|| ||::: | :|  ||:||:||:
Db     927 AATCTGCAGCATGACAGTGCAATAGAAAAGAAAAACCCATTTTCTGGGGAGAAATTCAAG 868

Qy     547 ytngcngcngarathtgyathtgyaaygargarytnaaygtnaaycccncargayaayggn 606
           :| |  || ||:||:||:||:  :||: |:||::  ||:|| ||:|  ||:||:||:||
Db     867 CTGGTTGCAGAAATTTGCATAAATAACAAGGAGCCAAATGTTAATCACCAAGACAACGGG 808

Qy     607 garaayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntgg 666
            |:||: |::   || |  |:||::|  : :     ||: :  |:  ::: :  ||  |
Db     807 AAAAATGTTTTCAGGGCACGTCAGAGACCCTTATGCCAGCCCCTCCCATCGCAGGCCCAG 748

Qy     667 mgnccnmgnmgnaa---rtggttytgyggnacnggnccnggnwsnytntgytgygtncar 723
           :| || :| :| ||   :|||||:||:||    || || ||  : :  ||:| :|| ||:
Db     747 AGGCCTAGGAGGAAAATATGGTTTTGTGGGCTGGGCCCAGGACCTTGCTGCTTTGTGCAG 688

Qy     724 ccnmgngayytngtnccntgygtnccngtnaaywsng---------------------- 760
```

```
           | :  ||::| || || ||: | ||       :
Db     687 TCTCAGGACTTGGTGCCCTGCATCCCATCTGTGGCTAAAAGGGGCCAATATACAGCTGAA 628

Qy     761 -cngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 819
            |  | || :: ||:|| |  :: || ||:|  |||  ::|  | :: || || | :||
Db     627 ACCATTGCTTCAGAGGGTGTAAGCCCCAAGCATTGGAGGCTTACATGTGGTGTTGGACCT 568

Qy     820 gtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarath 879
           || ||  |  |:||::: :| | :||:|| |||||:|| |    ::| ||:||:| :||
Db     567 GTGGGTACACAGAAGTCAAGAACTGAGGTTTGGGAACCTCTGCCTAGATTTCAGAGGATG 508

Qy     880 tayggnaayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgn 939
           ||:|| ||: | ||||  | :| ||:| :| :|| |  || || |   : :: |||:|
Db     507 TATGGAAACACCTGGATATCCAGGCAGAGGTGTGCTGCAGGGGTAGAGCCCTCATGGAGA 448

Qy     940 acnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccn 999
           || :: || :| |   | ||:||:|| ||:|| || | |||:|| || ||::| || ||
Db     447 ACCTCTGCTAGGGCAATGCAGAAGGGAAATGTGGGGTTGGAGCCTCCACACAGAGTCCCC 388

Qy    1000 wsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraar 1059
           .  :: ||       :: ::  | || || :| :|  : || || :: :: :| :  ||:||
Db     387 ACTGG------GACTAGTGGAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAAT 334

Qy    1060 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgycar 1119
           |  :|  : || ||::: :|  | | :|| || | :||: : ||  |:|| ||:||:||
Db     333 GATAGACCAACTGACAGCTTGGATCGTGTACCTGGAAAAGCCACAAACACTCAATGCCAT 274

Qy    1120 ccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacn 1179
           || || ||:|| ||  | | || : : || || | :||: |    |  || ||::|  |
Db     273 CCTGTGAAAGCAGCCAGGAGGGGAGCTGTACCCTGCAAAGCCACAGGGGCAGAGCTTCCC 214

Qy    1180 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1239
           ||: | || |  | : |::| :| ||:   | :| ::| ||:|| :| | :|| || ||:
Db     213 AAGGCTGTAGAAACCCACCTCTTACATCAGCGTGTCCTGGATGTGAGACCTGGAGTCAAA 154

Qy    1240 mgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggn 1299
           | ||:||:||:|| |  :| :| ||:| :||:|| || || ||::| |  | :|||||
Db     153 GGAGATCATTTTGGAGTTTTAAGATTTGGCTGCCCCACTGGATTTAGGAGTTGCATGGGG 94

Qy    1300 ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg 1358
           || ||  || :| ||:||:|| ||:||: |:|| | :|| |     || ||:|  ||:||
Db      93 CCTGTAGCCCCTTTGTTTTGGCCAATTTCTCCCATCTGGAATGAGTGTGTTTATCCAATG 34

Qy    1359 yytntayytncaytgyat 1376
           ::  ||::  ||:| :||
Db      33 CCCGTACCCCCATTATAT 16
```

Search completed: July 8, 2002, 16:20:51
Job time: 12424 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 12:57:56 ; Search time 104.62 Seconds
(without alignments)
3233.008 Million cell updates/sec

Title: US-09-997-610-3
Perfect score: 1377
Sequence: 1 athgtngtnathccngtnyt.....gyytntayytncaytgyatg 1377

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	299.2	21.7	49136	4	US-09-422-869-1	Sequence 1, Appli
c	2	177.6	12.9	14855	2	US-08-687-080-59	Sequence 59, Appl
	3	164.6	12.0	620	4	US-08-991-789A-29	Sequence 29, Appl
	4	164.6	12.0	620	4	US-09-062-451-29	Sequence 29, Appl
	5	140.6	10.2	152331	3	US-09-128-155-16	Sequence 16, Appl
	6	135.8	9.9	14855	2	US-08-687-080-59	Sequence 59, Appl
c	7	67	4.9	7218	1	US-08-232-463-14	Sequence 14, Appl
	8	59	4.3	2067	4	US-09-106-194-11	Sequence 11, Appl
	9	51.8	3.8	1313	2	US-08-463-911-6	Sequence 6, Appli
	10	51.8	3.8	4517	4	US-09-140-804-9	Sequence 9, Appli
c	11	50.6	3.7	1155	3	US-09-053-866-3	Sequence 3, Appli
	12	48.8	3.5	2082	4	US-09-440-325A-2	Sequence 2, Appli
	13	48.8	3.5	2088	4	US-09-351-414-3	Sequence 3, Appli
	14	48.6	3.5	289	4	US-09-007-005-17	Sequence 17, Appl
	15	48.6	3.5	289	4	US-09-244-796-17	Sequence 17, Appl
	16	48.6	3.5	1107	3	US-09-188-930-217	Sequence 217, App
	17	48.6	3.5	1276	2	US-08-463-911-1	Sequence 1, Appli
	18	48.2	3.5	2277	1	US-08-676-967-2	Sequence 2, Appli
	19	48.2	3.5	2277	1	US-08-676-974-2	Sequence 2, Appli
	20	48.2	3.5	2277	2	US-09-098-487-2	Sequence 2, Appli
c	21	48	3.5	2277	1	US-08-676-967-2	Sequence 2, Appli
c	22	48	3.5	2277	1	US-08-676-974-2	Sequence 2, Appli
c	23	48	3.5	2277	2	US-09-098-487-2	Sequence 2, Appli
c	24	44.4	3.2	1212	2	US-09-092-770-18	Sequence 18, Appl
c	25	44.4	3.2	1212	4	US-09-222-851-18	Sequence 18, Appl
	26	43.4	3.2	5181	1	US-08-257-073-10	Sequence 10, Appl
c	27	43	3.1	1215	2	US-09-092-770-8	Sequence 8, Appli
c	28	43	3.1	1215	4	US-09-222-851-8	Sequence 8, Appli
	29	42.4	3.1	1015	3	US-09-188-930-30	Sequence 30, Appl
c	30	42	3.1	2067	4	US-09-106-194-11	Sequence 11, Appl
	31	41.8	3.0	1001	3	US-09-188-930-218	Sequence 218, App
	32	41.8	3.0	2265	3	US-09-369-618-3	Sequence 3, Appli
	33	41.8	3.0	2265	3	US-09-369-617-3	Sequence 3, Appli
	34	41.6	3.0	510	2	US-08-934-959-7	Sequence 7, Appli
c	35	41	3.0	597	4	US-09-528-760A-3	Sequence 3, Appli
c	36	40.8	3.0	723	3	US-08-911-423-5	Sequence 5, Appli
	37	40.8	3.0	1659	1	US-08-231-729B-2	Sequence 2, Appli
	38	40.6	2.9	2949	4	US-09-412-554A-3	Sequence 3, Appli
	39	40.2	2.9	500	3	US-09-141-000-2	Sequence 2, Appli
	40	40.2	2.9	1659	1	US-08-231-729B-1	Sequence 1, Appli
c	41	40	2.9	2088	4	US-09-351-414-3	Sequence 3, Appli
c	42	39.8	2.9	2265	3	US-09-369-618-3	Sequence 3, Appli
c	43	39.8	2.9	2265	3	US-09-369-617-3	Sequence 3, Appli
	44	39.6	2.9	4379	1	US-08-592-214A-17	Sequence 17, Appl
	45	39.6	2.9	4379	3	US-09-149-976-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-422-869-1
; Sequence 1, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 49136
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-1

Query Match 21.7%; Score 299.2; DB 4; Length 49136;
Best Local Similarity 41.1%; Pred. No. 1.1e-72;
Matches 388; Conservative 186; Mismatches 307; Indels 62; Gaps 5;

```
Qy    444 rgcnttywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaa 503
          :|   |::: ||:| ::| ||:|| ||:||:||::: :  ||:||::| ||:|| ||: |
Db  47384 ggaactcacatttaaaagggaagcagagcataaaagtttagaaaatttgcatcctgacta 47443

Qy    504 ygtnathaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtg 563
          :|  ||: |:||:||:||: | ||::: ||:|  ||:||:||::  || || ||:||:||
Db  47444 tgggatagaaaagaaaaactcattttctgaggagaaattcaagccagctgcagaaatttg 47503

Qy    564 yathtgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggac 623
          :||: |:||: |:||::   |:|| ||:     :||:||:|  | :||: |:::  || |
Db  47504 cataagtaactaggagccacatgttaatagcatagacaatgagggaaatgtctccagggc 47563

Qy    624 ntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naa 680
           ||:|||:| :: :: ||:||: :   :  ::: :  || |||:| |  :| :|    ||
Db  47564 atgtcagaggtcttcacagcaaccccacccatcacaggcctggaggcttaggaggaaaaa 47623

Qy    681 rtggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtncc 740
```

```
Db  47624 atggttttgtcgttggggcccagggccttgctggtttgtgcagtctcaggacttggtgcc 47683

Qy    741 ntgygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggcaryt 800
Db  47684 ccacatcc------------cagcagtggctaaaaggggccaatgtacagcttagacctt 47731

Qy    801 nccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarccncc 860
Db  47732 tgcttcagaggtgcaagc----------------------------------------cc 47751

Qy    861 nathmgnttycaraarathtayggnaaycctggatgccnmgncaraarttygcngtngg 920
Db  47752 caagccttggtggcttacatgtggtgttgggcctgcagatacacagaagtttgctgcact 47811

Qy    921 ngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntggga 980
Db  47812 ggtggaaccctcatgtagaacctctgctagggcagtgtagaagtgatatgtggggttgga 47871

Qy    981 rccnccncaymgngtnccnwsnggngcnc------cnwsnwsnmgngcngtnmgnmgnws 1034
Db  47872 gccccccacacaatccccactggggcactgcctactgctactggaactgtgagaagaag 47931

Qy   1035 nccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccnga 1094
Db  47932 gccaccatcctccagaccccagaatggtagatccactgatggcttgaaccatgcacctgg 47991

Qy   1095 raarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccnta 1154
Db  47992 aaaagccacagacactcaacaccagcctgtgaaggcagctggaagggaggctgtaccctg 48051

Qy   1155 yaaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycayga 1214
Db  48052 caaaacaacagaggcagagctgcccaaggtcatgggagcccacctcttgcatgagcctga 48111

Qy   1215 yytngaygtnmgncayggngtnaarmgngaycayctttyggngcnytnmgnttygaytgycc 1274
Db  48112 cttgaatgtgagacatggagtcaaaggagatcattttggagctttaagttttgactgccc 48171

Qy   1275 nacnggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnt 1333
Db  48172 acctggatttcggacttgcatggggcctgtggccccttcattttggccaatttatcccat 48231

Qy   1334 tyggnacngcngtnttyacncartgyytntayytncaytgyat 1376
Db  48232 ttggaatgggtatatttacccaatgcctgtacccccattctat 48274
```

```
RESULT   2
US-08-687-080-59/c
; Sequence 59, Application US/08687080
; Patent No. 5965427
;  GENERAL 1NFORMAT1ON:
;    APPL1CANT:  Gregory Dolganov
;    T1TLE OF 1NVENT1ON:  Human RAD50 Gene and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  175
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      C1TY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      Z1P:  94306
;    COMPUTER READABLE FORM:
;      MED1UM TYPE:  Floppy disk
;      COMPUTER:  1BM PC compatible
;      OPERAT1NG SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patent1n Release #1.0, Version #1.25
;    CURRENT APPLICAT1ON DATA:
;      APPL1CAT1ON NUMBER:  US/08/687,080
;      FIL1NG DATE:  17-JUL-1996
;      CLASS1F1CAT1ON:  435
;    PRIOR APPL1CAT1ON DATA:
;      APPL1CAT1ON NUMBER:  US 08/592,126
;      FILING DATE:  26-JAN-1996
;    ATTORNEY/AGENT 1NFORMAT1ON:
;      NAME:  Sholtz, Charles K.
;      REG1STRAT1ON NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  4600-0111.30
;    TELECOMMUN1CAT1ON 1NFORMAT1ON:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  1NFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERIST1CS:
;      LENGTH:  14855 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHET1CAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      1ND1V1DUAL 1SOLATE:  5' END OF INTRON 2 OF RAD50 GENOM1C
;      1ND1V1DUAL 1SOLATE:  SEQUENCE
US-08-687-080-59


  Query Match             12.9%;  Score 177.6;  DB 2;  Length 14855;
  Best Local Similarity   44.4%;  Pred. No. 7.7e-39;
  Matches  221;  Conservative   90;  Mismatches  172;  1ndels   15;  Gaps    4;

Qy    890 cntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnm 949
Db  14854 CCTGGATGCCCAGGCAGATGTTTGCTGCAGGGGTGGGGCCCTTATGGAAAACCTCTGCTA 14795

Qy    950 gngtngtncaraarggnaaygtnggntgggarccncccncaymgngtnccnwsnggngcnc 1009
Db  14794 GGGCAATATGGAAGGGAAATGTGGGGTTGAAACCCCAC--AGAGTTCCTATGGAGGGGAC 14737

Qy   1010 cnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnyt----------ncaraar 1059
Db  14736 TGCCTAGTGGAGCTGTGAGAAGACAGCCACTGTCCTCCAGACTGGTAGATCCCCCAGAAT 14677

Qy   1060 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgycar 1119
Db  14676 AATAGATCCACTGACAGCTTGCACTGTGCACCTGGAAAAACTGCAGGCACTCAACACCAG 14617

Qy   1120 ccngtnaargcngcngggnatggarwsngtnccntayaaracngtngtngcngarytnacn 1179
Db  14616 CCTGTGAAAACAGCCAGGAAGGAGGCTATACCCTGCAAAGCC--AGAAGTGGAGCTGCCC 14559

Qy   1180 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1239
Db  14558 AAGGCCATGGAAGCCCACCTCTTGCATCAGAGTGACCTGGATGTGAGACATGGAGTCAAA 14499

Qy   1240 mgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggn 1299
Db  14498 GGAGATCATTCTGGAGCTTTAAGATACACCTGCCCCACTGAATTTCGGACTTGCACGGGG 14439

Qy   1300 ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg 1358
Db  14438 CCTGTAGCCCCTTTGTTTTGGCCAATTTCTCCCATTTGGAATGGCTGTATTTGCCCAATG 14379

Qy   1359 yytntayytncaytgyat 1376
Db  14378 CCTGTATCCCCATTGTAT 14361
```

```
RESULT   3
US-08-991-789A-29
; Sequence 29, Application US/08991789A
; Patent No. 6225054
;   GENERAL 1NFORMAT1ON:
;        APPL1CANT: Frudakis, Tony N.
;                   Smith, John M.
;                   Reed, Steven G.
```

```
;       TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;                           TREATMENT AND DIAGNOSIS OF BREAST CANCER
;       NUMBER OF SEQUENCES: 292
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Seed IP Law Group
;            STREET: 701 Fifth Avenue, Suite 6300
;            CITY: Seattle
;            STATE: Washington
;            COUNTRY: USA
;            ZIP: 98104-7092
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/991,789A
;            FILING DATE: 11-Dec-1997
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Potter, Jane E. R.
;            REGISTRATION NUMBER: 33,332
;            REFERENCE/DOCKET NUMBER: 210121.419C3
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206) 622-4900
;            TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 29:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 620 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-08-991-789A-29


  Query Match             12.0%;  Score 164.6;  DB 4;  Length 620;
  Best Local Similarity   40.7%;  Pred. No. 1.6e-36;
  Matches  250;  Conservative  105;  Mismatches  243;  Indels   17;  Gaps    6;

Qy    772 garggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaar 831
          ||:||  | :: || ||:|| || ||::| || :  || ||  |:|| || || ||  |:
Db      3 GAGGGTACCAGCCCCAAGCCTTGACAACTTCCATAGGGTGTCAAGCCTGTGGGTGCACAG 62

Qy    832 aarwsnmgnathgargtntgggarccnccnathmgnttycaraarathtaygg------- 884
          ||::: :  ||:||: | ||||| || |    ::| ||:||:| :||:||  |
Db     63 AAGTCAAAAATTGAGTTTTGGGATCCTCAGCCTAGATTTCAGAGGATATAAAGAAACACC 122

Qy    885 naayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnws 944
           ||: | | |||    :| ||:||:||: |    ||  |       ::  ||:  || ::
Db    123 TAACACCTAGATATTCAGACAAAAGTTTACTACAGGGATGAAGCTTTCACGGAAAACCTC 182

Qy    945 ngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsngg 1004
            | :|    || ||:||: | ||:|| || | |||:|| ||  |::|  | || ::  |
Db    183 TACTAGGAAAGTACAGAAGAGAAATGTGGGTTTGGAGCCCCCAAACAGAATCCCCTCTAG 242

Qy   1005 ngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmg 1064
            | |   : :      | || :| :| :: || |  :: :: :| :  ||:|| |  :|
Db    243 AACACTGCCTAATGAAACTGTGAGAAGATGGCCACTGTCATCCAGACACCAGAATGATAG 302

Qy   1065 nwsnacngaywsnytncarcayg-tnccngaraarwsnacngayacncartgycarcc-- 1121
             : ||  | :  :|    ||:  | ||   :||: : || ||:|| ||:||:||:||
Db    303 ACCCACCAAAAACTTATGCCATATTGCCTATAAAACCTACAGACACTCAATGCCAGCCCC 362

Qy   1122 -ngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacna 1180
             |  ||:    | |  | | |::: ||||| ||:||  |    |  || ||::|  |
Db    363 ATGAAAAAAAAACTGAGAAGAAGACTGTNCCCTACAATGCCACCGGAGCAGAACTGCCCC 422

Qy   1181 aracngtnggnathtayytnytncaytgycaygayytngayg-tnmgncayggngtnaar 1239
           |: |  | |    : | :| :|  |:     :||::| ||:| | :| ||:||  |   :
Db    423 AGGCCATGGAAGCACAGCTCTTATATCAATGTGACCTGGATGTTGAGACATGGAATCCNA 482

Qy   1240 mgngaycayttyggngcnytnmgntt----ygaytgyccnacnggnttymgnacntayat 1295
```

```
          |  |:| :||:  | | :  :   |    :||:||:|| |     ||: | || || ||
Db    483 NGAAATCNTTTTAANACTTCCACGGTTNAATGACTGCCCTATTANATTCNGAACTTANAT 542

Qy   1296 g-ggnccngtnccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacnc 1354
            | || ||  | :  |   |: | | :||:  :|  ||:|| |  ||  | ||:   |
Db    543 CCNGGCCTGTGACCTCTTTGCTTTGGCCATTCCCCCTTTTTGGAATGGCTNTTTTTTTCC 602

Qy   1355 artgyytntayytnc 1369
             :||::| | ::  |
Db    603 CATGCCTGTNCCCTC 617


RESULT   4
US-09-062-451-29
; Sequence 29, Application US/09062451
; Patent No. 6344550
;  GENERAL INFORMATION:
;    APPLICANT:  Frudakis, Tony N.
;    APPLICANT:  Smith, John M.
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE
;    TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF BREAST CANCER
;    NUMBER OF SEQUENCES:  297
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/062,451
;      FILING DATE:  04-APR-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.419C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  620 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-062-451-29


  Query Match             12.0%;  Score 164.6;  DB 4;  Length 620;
  Best Local Similarity   40.7%;  Pred. No. 1.6e-36;
  Matches  250;  Conservative  105;  Mismatches  243;  Indels   17;  Gaps    6;

Qy    772 garggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaar 831
          ||:||  | :: || ||:|| || ||::| || :  || ||  |:|| || || ||  |:
Db      3 GAGGGTACCAGCCCCAAGCCTTGACAACTTCCATAGGGTGTCAAGCCTGTGGGTGCACAG 62

Qy    832 aarwsnmgnathgargtntgggarccnccnathmgnttycaraarathtaygg------- 884
          ||::: :  ||:||: | ||||| || |    ::| ||:||:| :||:||  |
Db     63 AAGTCAAAAATTGAGTTTTGGGATCCTCAGCCTAGATTTCAGAGGATATAAAGAAACACC 122

Qy    885 naayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnws 944
           ||: | | |||    :| ||:||:||: |    ||  |       ::  ||:  || ::
Db    123 TAACACCTAGATATTCAGACAAAAGTTTACTACAGGGATGAAGCTTTCACGGAAAACCTC 182

Qy    945 ngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsngg 1004
            | :|    || ||:||: | ||:|| || | |||:|| ||  |::|  | || ::  |
```

```
Db    183 TACTAGGAAAGTACAGAAGAGAAATGTGGGTTTGGAGCCCCCAAACAGAATCCCCTCTAG 242

Qy   1005 ngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmg 1064
          | |   : :      | || :| :| :: || |  :: :: :| :  ||:|| |  :|
Db    243 AACACTGCCTAATGAAACTGTGAGAAGATGGCCACTGTCATCCAGACACCAGAATGATAG 302

Qy   1065 nwsnacngaywsnytncarcayg-tnccngaraarwsnacngayacncartgycarcc-- 1121
          : ||  | :  :|    ||:  | ||   :||: : || ||:|| ||:||:||:||
Db    303 ACCCACCAAAAACTTATGCCATATTGCCTATAAAACCTACAGACACTCAATGCCAGCCCC 362

Qy   1122 -ngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacna 1180
           |  ||:    | |  | | |::: ||||| ||:||  |    |  || ||::|  |
Db    363 ATGAAAAAAAAACTGAGAAGAAGACTGTNCCCTACAATGCCACCGGAGCAGAACTGCCCC 422

Qy   1181 aracngtnggnathtayytnytncaytgycaygayytngayg-tnmgncayggngtnaar 1239
          |: |  | |    : | :| :|  |:    :||::| ||:| | :| ||:||  |   :
Db    423 AGGCCATGGAAGCACAGCTCTTATATCAATGTGACCTGGATGTTGAGACATGGAATCCNA 482

Qy   1240 mgngaycayttyggngcnytnmgntt----ygaytgyccnacnggnttymgnacntayat 1295
           |  |:| :||:  | | :  :   |    :||:||:|| |     ||: | || || ||
Db    483 NGAAATCNTTTTAANACTTCCACGGTTNAATGACTGCCCTATTANATTCNGAACTTANAT 542

Qy   1296 g-ggnccngtnccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacnc 1354
           | || ||  | :  |   |: | | :||:  :|  ||:|| |  ||  | ||:   |
Db    543 CCNGGCCTGTGACCTCTTTGCTTTGGCCATTCCCCCTTTTTGGAATGGCTNTTTTTTTCC 602

Qy   1355 artgyytntayytnc 1369
           :||::| | ::  |
Db    603 CATGCCTGTNCCCTC 617


RESULT   5
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
;   LENGTH: 152331
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(152331)
;   OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16


  Query Match           10.2%;  Score 140.6;  DB 3;  Length 152331;
  Best Local Similarity  32.8%;  Pred. No. 2.1e-27;
  Matches  324;  Conservative  168;  Mismatches  426;  Indels   71;  Gaps    6;

Qy    417 hcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngarcayaa 476
          :||:| :  |    |  :|     :|    ::  ||:| ::  ||:|  ||:  :||
Db  71225 tcagggtaaagaaatgacttaaagttggaacctatgtttaaaatggaagtagagtctaa 71284

Qy    477 rwsnwsngaraayytncaycengayaaygtnathaaraaraaraaycenttywsngargg 536
          ::  :  ||:||::| || || | :  :||   : |:||: |:  :   |   : | :|
Db  71285 aaatttggaaaatttgcagcctggccttgtggcagagaaagaatccaagtaggctgcaga 71344

Qy    537 naarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaayccnca 596
            |:|    ::| |     ||:||: |:||   :||   :||::  | :|  ||:   ||

Db  71345 gcaatcattgctag----agagattagcatgactaaaagggagccaagtgctaatattca 71400

Qy    597 rgayaaygngngaraayathwsntggacntgycarmgnwsnwsncarcarwsnathaarws 656
          :||:||:|   |:||   ::   || | | :||::| :: :   |:||: :  |:  :::
Db  71401 agacaatgttaaaaaggccttgagggcatttcagagatctatgaagcagcccctcccatc 71460

Qy    657 nytngcntggmg-------nccnmgnmgnaartggttytgyggnacnggnccnggnwsny 709
           :  |    |:|        |         |: ||||:| :||  |    || ||  : :
Db  71461 acaggtgcagaggtttggtgcactaggcccagaggttttatgggccanngccagggccac 71520

Qy    710 tntgytgygtncarccnmgngayytngtnccntgygtnccngtnaa-------------- 755
            ||:| :|  ||:|   | ||:   |   |  |: | |  |  |
Db  71521 actgctatgcacagctttgggacactgctgcccgcatccaggccactctgctctggctcc 71580

Qy    756 ----------------------------ywsngcngtngcnwsngarggngcnwsncc 785
                                       :::  |   ||  : ||:||  | :: |
Db  71581 acccttggctcaaacgggccaagatagagcttggaccactgctcccgagggcacaagcca 71640

Qy    786 naarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnath-- 843
           ||:|| |||  ::| || :: || ||  |:|| |  || ||  |:||  : :|  |
Db  71641 taagccttggtggtttccatgtggtgttaagcctgcaggtgcccagaatgcaagattgag 71700

Qy    844 gargtntgggarccnccnathmgnttycaraarathtayggnaayccntggatgccnmgn 903
          | :|  |||| :|  ||   ::  ||:||:| :|| | :   || || | | | || :|
Db  71701 ggagcttgggcacttccacctaaatttcagaggatgtgtcagaaaccctaggttcccagg 71760

Qy    904 caraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraar 963
          ||:||:  :|     || |  |   : :    |:| || ::  | :| |   | | :||:
Db  71761 cagaagcatgatacaggggcagagcccttgcagagaacctctactagggcaatgccaaag 71820

Qy    964 ggnaaygtnggntgggarccnccncaymgngtnccnwsnggngncnccnwsnwsnmgngcn 1023
          |  ||:|| || | |||: |  | ||::  || || :: || || |   :  :     |
Db  71821 gaaaatgtggggttggagtcctcacacatggtccccactggggcactacctggtgatact 71880

Qy   1024 gtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncar 1083
          ||  | :   : |  |   : :: :| :  ||:|| || :| :: || | ::: :  ||
Db  71881 gtgggaatggggctgctgcccctccagaccccagaatggtagatgcactggcagctggcac 71940

Qy   1084 caygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggnatggar 1143
          | :|  || | :||: :  | | :|| ||:  :||:||  | | : | ||     ||
Db  71941 cctgagcctggaaaagctgcaggcactcaactccaacccatgagatcagccacatgggct 72000

Qy   1144 wsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtayytnytn 1203
          ::  | ||   ||: |   |  || | ::|  | ||: |  | ||   :||::  :|
Db  72001 ac--tcccagggaagcccacagaggcagggctgtctaaggccttgggagcctacccctg 72058

Qy   1204 caytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgn 1263
           |:   | :|      |:|  :   |             ||: |: |:|  || :| :|
Db  72059 aaccagcttgcaggacatggaatcaa-----------agattatgttgcagctttaagg 72106

Qy   1264 ttygaytgyccnacnggnttymgnacntayatgggncengtnccnytntgyttyggncar 1323
           |: |:  :    | |    :   |    : || |     |    :| | :||:
Db  72107 cttaatgttttccctgtcaatttcaggcttgtgtgggacctgttgctttttttttttttt 72166

Qy   1324 ttyttyccnttyggnacngcngtnttyac 1352
          ||:||:   || |  || |  || |  ||
Db  72167 ttttttttttggtcacaggtgtttgaac 72195


RESULT   6
US-08-687-080-59
; Sequence 59, Application US/08687080
; Patent No. 5965427
;  GENERAL INFORMATION:
;    APPLICANT:  Gregory Dolganov
;    TITLE OF INVENTION:  Human RAD50 Gene and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  175
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
```

```
;       ZIP:  94306
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/687,080
;       FILING DATE:  17-JUL-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/592,126
;       FILING DATE:  26-JAN-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Sholtz, Charles K.
;       REGISTRATION NUMBER:  38,615
;       REFERENCE/DOCKET NUMBER:  4600-0111.30
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 324-0880
;       TELEFAX:  (415) 324-0960
;   INFORMATION FOR SEQ ID NO:  59:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  14855 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       INDIVIDUAL ISOLATE:  5' END OF INTRON 2 OF RAD50 GENOMIC
;       INDIVIDUAL ISOLATE:  SEQUENCE
US-08-687-080-59


  Query Match             9.9%;  Score 135.8;  DB 2;  Length 14855;
  Best Local Similarity   39.5%;  Pred. No. 4.9e-27;
  Matches  191;  Conservative  101;  Mismatches  160;  Indels   31;  Gaps    3;

Qy    444 rgcnttywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaa 503
          :|   |::  ||:| ::| ||:|| ||:||:||::: :  ||:||::| ||:|| ||: |
Db   5187 GGAACTTATATTTAAAAGGGAAGCAGAGCATAAAAGTTTGGAAAATTTGCACCCTGATCA 5246

Qy    504 ygtnathaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtg 563
          :|| |: |:||:||:||:|| ||::: | :|   |:||:||::  || || ||:||:||
Db   5247 TGTGGTAGAAAAGAAAAACCCATTTTCTGGGGAGGAATTCAAGCCGGCTGCAGAAATTTG 5306

Qy    564 yathtgyaaygargarytnaayg-tnaayccncargayaayggngaraayathwsntgga 622
          :||: |:|| ||:||::  ||:| | ||:   ||:||:| :|| ||:||:  ::    |
Db   5307 CATAAGTAAAGAGGAGCAGAATGTTTAATAGCCAAGACACTGGGGAAAATGGCTAGAAGG 5366

Qy    623 cntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaart 682
          | | :||::| :: :: ||:||: :  |:  ::  :  |     :  ||
Db   5367 CATTTCAGAGATCTTCACAGCAGCCCCTCCCATTACAGGACACTACTCCC---------- 5416

Qy    683 ggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccnt 742
          || :| : |          :: :   |:|        || :|  |: |
Db   5417 TGTGTTTCAGCCACTCCAGCTCCAGCCATGGCTAAAAGGGCCCCAGATATATC------- 5469

Qy    743 gygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggcarytnc 802
                  ::  || |  ||  : ||:|  || ::    ||:|| |||||::| |
Db   5470 ------------TTAAGCTGCTGCTCCAGAGGCTGCAAGCTGTAAGCCTTGGCAGCTTC 5516

Qy    803 cnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarccnccna 862
          |    || ||  |:|| |  |  ||  |:| ::: :|  |:||:|  |||||:|| ||
Db   5517 CACATGGTGTTAAGCCTGGGGTTGCGCAGAGAACAAGAGTTGAGGCTTGGGAGCCTCCAC 5576

Qy    863 thmgnttycaraarathtayggnaayccntggatgccnmgncaraarttygcngtnggng 922
          ::| ||:||:  :|| ||:|| ||:   |||||| | :| ||:||:| :|| |     |
Db   5577 CTAGATTTCAGGGGATGTATGGAAACGTCTGGATGTCCAGGCAGAAGTCTGCAGGACTTG 5636

Qy    923 tng 925
          | |
Db   5637 TTG 5639


RESULT   7
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             4.9%;  Score 67;  DB 1;  Length 7218;
  Best Local Similarity   12.6%;  Pred. No. 6.4e-08;
  Matches   64;  Conservative  203;  Mismatches  241;  Indels    0;  Gaps    0;

Qy    226 athahttyacnggngtnytntayaaygcncarmgngayytnaargargcnatgggngtn 285
          ||:  :| :|   |  | : |    : |  |::  | : |  |:| : |  ||    |
Db   1536 ATCGCTTCTAGACGCATCTATTCAGTTTCAAAAAACGGCATGTAGGCATCACTGTAATTA 1477

Qy    286 ttygcntgymgngtnccnggnaaytaytaywsnwsnttygaygtngarytncaycaytgy 345
            :   | :              |   |:   :: : |   |::  ::|    :  :  :
Db   1476 CCTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRR 1417

Qy    346 aargtnaayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaar 405
          ::|:  :: : : ::   : ::: ::| :|: :   :  :: ::|::|::|: ::: ::|
Db   1416 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1357

Qy    406 carcarwsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmgngar 465
            :| :|:: : : :|::|:  :   :::       ::|:     ::    : ::: ::|
Db   1356 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1297
```

```
Qy    466 gcngarcayaarwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccn 525
Db   1296 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1237

Qy    526 ttywsngarggnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaay 585
Db   1236 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1177

Qy    586 gtnaayccncargayaayggngaraayathwsntggacntgycarmgnwsnwsncarcar 645
Db   1176 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1117

Qy    646 wsnathaarwsnytngcntggmgnccnmgnmgnaartggttytgyggnacnggnccnggn 705
Db   1116 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCT 1057

Qy    706 wsnytntgytgygtncarccnmgngayy 733
Db   1056 CCCTCGACCTGCAGCCAAGCTCGGAATT 1029


RESULT   8
US-09-106-194-11
; Sequence 11, Application US/09106194
; Patent No. 6262234
;  GENERAL INFORMATION:
;    APPLICANT:  Holloway, James
;    APPLICANT:  Jelinek, Laura
;    APPLICANT:  Durnam, Diane
;    APPLICANT:  Blumberg, Hal
;    TITLE OF INVENTION:  NOVEL NUCLEAR RECEPTOR POLYPEPTIDE
;    TITLE OF INVENTION:  ZPPAR4
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/106,194
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leith, Debra K
;      REGISTRATION NUMBER:  32,619
;      REFERENCE/DOCKET NUMBER:  96-11
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6674
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2067 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-106-194-11


  Query Match             4.3%;  Score 59;  DB 4;  Length 2067;
  Best Local Similarity   9.3%;  Pred. No. 3.5e-06;
  Matches  101;  Conservative  297;  Mismatches  676;  Indels   12;  Gaps    3;

Qy     71 cnmgnccnccngargargtnggnccnccnggngcnccnggnytnccncartayacnggng 130
Db     13 YTNGARMGNCCNCCNGGNCARGTNWSNYTNGCNGCNCCNTTYYTNCCNGTNACNGARCAR 72

Qy    131 arathwsngaratgacnaartgyccntgyccngayathgarmgnwsngcnttyacngtna 190
Db     73 AARTTYACNYTNTGGATHMGNCARGCNCARATGWSNTGYGARAAYGCNCAYGCNYTNCCN 132

Qy    191 arytnwsnggnaarytnccnytnccnttyaarccnathathttyacnggngtnytntaya 250
Db    133 WSNACNCCNCCNTTYTTYYTNTTYTTYYTNTTYCCNTTYTTYCCNYTNG------TNCCN 186

Qy    251 aygcncarmgngayytnaargargcnatgggngtnttygcntgymgngtnccnggnaayt 310
Db    187 WSNCCNWSNWSNYTNYTNACNAARCCNCCNCCNACNCCNCCNCCNGCNGCNGCNYTNCCN 246

Qy    311 aytaywsnwsnttygaygtngarytncaycaytgyaargtnaayathtggytnatgmgna 370
Db    247 CCNCCNCCNCARWSNAARWSNCARAARAARWSNWSNGARCCNGARGCNGTNGCNTTYTTY 306

Qy    371 arcarathytngcnaayaargargarathwsnaarcarcarwsnathcargargtnacnt 430
Db    307 AARAARCARGCNCAYTGGMGNGARMGNAARMGNAARACNAARCCNAARCARAAYCCNGGN 366

Qy    431 gggtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnwsngaraayy 490
Db    367 ACNMGNCARCCNGARCAYTTYTTYTTYACNYTNCCNGARAAYAARCARACNAAYAARCAR 426

Qy    491 tncayccngayaaygtnathaaraaraaraayccnttywsngarggnaarttyaarytng 550
Db    427 WSNWSNAARCARWSNCCNCCNACNWSNAARYTNYTNACNMGNMGNMGNMGNCARACNWSN 486

Qy    551 cngcngarathtgyathtgyaaygargarytnaaygtnaayccncargayaayggngara 610
Db    487 CCNTGYWSNCAYGGNGTNMGNYTNAARGGNTGGTTYWSNMGNCARWSNWSNWSNWSNCCN 546

Qy    611 ayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnc 670
Db    547 ACNACNTTYTTYACNMGNGCNGARMGNGAYTTYTGGGCNYTNMGNGGNWSNGGNTGGGAR 606

Qy    671 cnmgnmgnaartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgng 730
Db    607 CARYTNCAYGAYTAYGCNGARMGNGARWSNGGNCAYACNATGMGNGCNCARATHGARGTN 666

Qy    731 ayytngtnccntgygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarc 790
Db    667 ATHCCNTGYAARATHTGYGGNGAYAARWSNWSNGGNATHCAYTAYG---GNGTNATHACN 723

Qy    791 cntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtnt 850
Db    724 TGYGARGGNTGYAARGGNTTYTTYMGNMGNWSNCARCARAAYAAYGCNWSNTAYWSNTGY 783

Qy    851 gggarccnccnathmgnttycaraarathtayggnaayccntggatgccnmgncaraart 910
Db    784 CCNMGNCARMGNAAYTGYYTNATHGAYGGNACNAAYMGNAAYMGNTGYCARCAYTGYMGN 843

Qy    911 tygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaayg 970
Db    844 YTNCARAARTGYYTNGCNYTNGGNATGWSNMGNGAYGCNGTNAARTTYGGNMG---NATG 900

Qy    971 tnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnm 1030
Db    901 WSNAARAARCARMGNGAYWSNYTNTAYGCNGARGTNCARAARCAYCARCARMGNYTNCAR 960

Qy   1031 gnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnc 1090
Db    961 GARCARMGNCARCARCARWSNGGNGARGCNGARGCNYTNGCNMGNGTNTAYWSNWSNWSN 1020

Qy   1091 cngaraarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnc 1150
Db   1021 ATHWSNAAYGGNYTNWSNAAYYTNAAYAAYGARACNWSNGGNACNTAYGCNAAYGGNCAY 1080

Qy   1151 cntaya 1156
```

```
            |::
Db   1081 GTNATH 1086


RESULT   9
US-08-463-911-6
; Sequence 6, Application US/08463911
; Patent No. 5869330
;  GENERAL INFORMATION:
;    APPLICANT:  Scherer, Philipp E.
;    APPLICANT:  Lodish, Harvey F.
;    TITLE OF INVENTION:  A NOVEL SERUM PROTEIN PRODUCED
;    TITLE OF INVENTION:  EXCLUSIVELY IN ADIPOCYTES
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,911
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  WHI95-05
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1313 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  73..804
US-08-463-911-6


  Query Match             3.8%;  Score 51.8;  DB 2;  Length 1313;
  Best Local Similarity   34.2%;  Pred. No. 0.00024;
  Matches   75;  Conservative   36;  Mismatches  108;  Indels    0;  Gaps    0;

Qy     158 gyccngayathgarmgnwsngcnttyacngtnaarytnwsnggnaarytnccnytnccnt 217
           |: |  |: |: | :| :: || ||:|  ||   ::|        |  |  |  | ||
Db     392 GTGCCTATGTATACCGCTCAGCATTCAGTGTGGGATTGGAGACTTACGTTACTATCCCCA 451

Qy     218 tyaarccnathathttyacnggngtnytntayaaygcncarmgngayytnaargargcna 277
           :| :|| ||:  :||:||     | :| ||:||:   ||::   |::   | |   | |
Db     452 ACATGCCCATTCGCTTTACCAAGATCTTCTACAATCAGCAAAACCACTATGATGGCTCCA 511

Qy     278 tgggngtnttygcntgymgngtnccnggnaaytaytaywsnwsnttygaygtngarytnc 337
             ||    ||:   ||::   | || ||    ||:||::   : | : |: |   : |
Db     512 CTGGTAAATTCCACTGCAACATTCCTGGGCTGTACTACTTTGCCTACCACATCACAGTCT 571

Qy     338 aycaytgyaargtnaayathtggytnatgmgnaarcara 376
           |:       | || ||  |: | :|  | :  ||: | |
Db     572 ATATGAAGGATGTGAAGGTCAGCCTCTTCAAGAAGGACA 610


RESULT  10
US-09-140-804-9
; Sequence 9, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT: Humes, Jacqueline M.
;  TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
;  FILE REFERENCE: 97-49
;  CURRENT APPLICATION NUMBER: US/09/140,804
;  CURRENT FILING DATE: 1998-08-26
;  EARLIER APPLICATION NUMBER: 60/056,983
;  EARLIER FILING DATE: 1997-08-26
;  NUMBER OF SEQ ID NOS: 47
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
;   LENGTH: 4517
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-140-804-9


  Query Match             3.8%;  Score 51.8;  DB 4;  Length 4517;
  Best Local Similarity   34.2%;  Pred. No. 0.0008;
  Matches   75;  Conservative   36;  Mismatches  108;  Indels    0;  Gaps    0;

Qy     158 gyccngayathgarmgnwsngcnttyacngtnaarytnwsnggnaarytnccnytnccnt 217
           |: |  |: |: | :| :: || ||:|  ||   ::|        |  |  |  | ||
Db     346 gtgcctatgtataccgctcagcattcagtgtgggattggagacttacgttactatcccca 405

Qy     218 tyaarccnathathttyacnggngtnytntayaaygcncarmgngayytnaargargcna 277
           :| :|| ||:  :||:||     | :| ||:||:   ||::   |::   | |   | |
Db     406 acatgcccattcgctttaccaagatcttctacaatcagcaaaaccactatgatggctcca 465

Qy     278 tgggngtnttygcntgymgngtnccnggnaaytaytaywsnwsnttygaygtngarytnc 337
             ||    ||:   ||::   | || ||    ||:||::   : | : |: |   : |
Db     466 ctggtaaattccactgcaacattcctgggctgtactactttgcctaccacatcacagtct 525

Qy     338 aycaytgyaargtnaayathtggytnatgmgnaarcara 376
           |:       | || ||  |: | :|  | :  ||: | |
Db     526 atatgaaggatgtgaaggtcagcctcttcaagaaggaca 564


RESULT  11
US-09-053-866-3/c
; Sequence 3, Application US/09053866
; Patent No. 6111075
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Wenfeng
;    APPLICANT:  Presnell, Scott R.
;    APPLICANT:  Yee, David P.
;    APPLICANT:  Foster, Donald C.
;    TITLE OF INVENTION:  PROTEASE-ACTIVATED RECEPTOR
;    TITLE OF INVENTION:  PAR4 (ZCHEMR2)
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/053,866
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Leith, Debra K
;      REGISTRATION NUMBER:  32,619
;      REFERENCE/DOCKET NUMBER:  98-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6674
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1155 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-053-866-3


  Query Match             3.7%;  Score 50.6;  DB 3;  Length 1155;
  Best Local Similarity   6.3%;  Pred. No. 0.00046;
  Matches   64;  Conservative  256;  Mismatches  703;  Indels    0;  Gaps    0;

Qy    302 cnggnaaytaytaywsnwsnttygaygtngarytncaycaytgyaargtnaayathtggy 361
          : |: |    :     :   :: ::   ::: : |  | ::    :    |::|::|:
Db   1071 YTGRAANARNCCNGCNCKNACYTTRTCNCKRAAYTCNGCNSWNACRTARTARTADATRAA 1012

Qy    362 tnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcarwsnathcarg 421
          :   :  :: ::  :   :: :: :  : :|    :|:   :   :   :  ::
Db   1011 NGGRTCNACRCANSWRTTNARNGTNSWNARNGCNARNSWNGGNACRTANGCNCCRTANAR 952

Qy    422 argtnacntgggtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnw 481
          :     |   |   :     : :|    :|:   :  ::  : ::  : : ::  :|
Db    951 RTTNCCCCANGCNSWNGGNSWNGGRTCNSWRTARTGNARNARNARNARNARRTTNSWNGG 892

Qy    482 sngaraayytncayccngayaaygtnathaaraaraaraayccnttywsngarggnaart 541
            ::|:    |  |    ::     : :  :  :  :      :  :   : |   :
Db    891 NACRAARAANGCNACNGCNSWNGCNARNACNACNGCNGTNARNCKNARNGCRTGNCCRTA 832

Qy    542 tyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaayccncargaya 601
          :  :          :  :    :  :    :  :  |::   :   |   :   :
Db    831 NCKNCKNCCNSWNGCNGCNARNGTRTGNARNGTNGCNCCRTARCANARNARCATNGCNAR 772

Qy    602 ayggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytng 661
            :  | ::: | ::     :     | ::: :  ::   :| : :   :|:    :
Db    771 NARNGGNARRAARCANCCNARNARNGCNARRCANGTRAANGCNGGYTGCCARTGNSWNGC 712

Qy    662 cntggmgnccnmgnmgnaartggttytgyggnacnggnccnggnwsnytntgytgygtnc 721
          :   |:: |  :     :  :  :: :           |  :       ::
Db    711 YTGNGCRTCNARNGGNARNGCRTCRTGRCANARNACNCKRTCNSWNCKNGCNARNCKRAA 652

Qy    722 arccnmgngayytngtnccntgygtnccngtnaaywsngcngtngcnwsngarggngcnw 781
           : :   : : :          |          :          ::   :     | :
Db    651 NGTYTGNCKYTGNARNGTNARNGGNARNGCNARNGCNGCNGCCATNARCCANGCNGCCAT 592

Qy    782 snccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgna 841
          :    :  |    || :      :     :  : :|  :     :  : : :    :  :
Db    591 RCANARNCCNARNGCNARNCKNCKNCCNCKNARNGCNCKNGCNCKNARNGGRTGNACNAR 532

Qy    842 thgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggatgccnm 901
            ::     :: |     : ::     | :: :: :      : :|              :
Db    531 NGCNARRTANCKRTCNARNSWNACNGCNGCNARNARNARNACNSWNCCRTACATRTGNCC 472

Qy    902 gncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncara 961
          :  :: :          :  :: :      :    :  || :  :: |  |
Db    471 RTANARNGCNGCNGTNGCNARNCKRCANGCNGCYTCNCCRAANGGCCANCKYTGNCCNCK 412

Qy    962 arggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgng 1021
           :::  ::    : |  : :          :  :  |  :  :  :  |     :  :  :
Db    411 NARRTGRTANGCDATNCKNGGNGGNARNGCNARNGCNARNARNARRTCNGCNGTNGCNAR 352

Qy   1022 cngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytnc 1081
              :  ::       |    :  :    :  |          :    ::      |
Db    351 RTTCATNARNARCATNGTNSWNGGNARNCKNGGNGCYTGNGTNGCNARNACCCANARNGC 292

Qy   1082 arcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggnatgg 1141
```

```
          :  :        :  ::             :      |     : :   |     :  :
Db    291 NARNCCRTTNGCNGGNARNCCNACNACNARNACNARNCCRTANARNGCNGGNACNARNCK 232

Qy   1142 arwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtayytny 1201
          ::  |  ||   :  ::  :  :     : ::  :    :   ::    :  |::  :
Db    231 NGTNGGNACCCANCCNARNARNARNGCNCKNSWNSWRTCNGGNARYTCNARNGTRTCNSW 172

Qy   1202 tncaytgycaygayytngaygtnmgncayggngtnaarmgngaycaytttyggngcnytnm 1261
            |::|  |::   :: | :    :  :   :    ::  |   : | :   |  :  :
Db    171 RTCRTTNGCRCANACYTGNCCNGGRTANCCNCKNGGNGCNGGNARDATNSWNGGNGTNSW 112

Qy   1262 gnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytntgyttyggnc 1321
          :   :  :  :|       : ::  ::  ::       |         ::    | :   |
Db    111 RTCRTCNCCNCCNCCNGTNSWNCCNSWYTCRTCRTANACNSWNGGNGTYTGNGTNCCNCC 52

Qy   1322 art 1324
          ::
Db     51 NSW 49


RESULT  12
US-09-440-325A-2
; Sequence 2, Application US/09440325A
; Patent No. 6280994
; GENERAL INFORMATION:
;  APPLICANT:  Sheppard, Paul O.
;  TITLE OF INVENTION:  Zac1: A Human Metalloenzyme
;  FILE REFERENCE:  98-79
;  CURRENT APPLICATION NUMBER: US/09/440,325A
;  CURRENT FILING DATE:  1999-11-15
;  NUMBER OF SEQ ID NOS:  3
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2082
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: This degenerate sequence encodes the amino acid
;   OTHER INFORMATION: sequence of SEQ ID NO:1.
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(2082)
;   OTHER INFORMATION: n = A,T,C or G
US-09-440-325A-2


  Query Match             3.5%;  Score 48.8;  DB 4;  Length 2082;
  Best Local Similarity   10.4%;  Pred. No. 0.0026;
  Matches   82;  Conservative  191;  Mismatches  512;  Indels    0;  Gaps    0;

Qy    356 thtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcarwsna 415
          |::||       :|   ::   |:|  :   |:  ::   |:: :::   ::   |:::
Db    316 ttyggnacncargcncayytnttyaargtnacncarttyaargayccngaygtnaayggn 375

Qy    416 thcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngarcaya 475
            :    ::   ::    :  | :   |::    : :  ::  ::|   :::  ::   :
Db    376 atgytnwsnaarytncaraayathgayaargcngcnytnwsnaargaygarytnmgngar 435

Qy    476 arwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccnttywsngarg 535
            ::  :: :    :      :|:|   |: :: |          | ::  |:
Db    436 tayaaygarytnytngcnnnnytngaratgacntaywsnatggcncargtntgyytnaay 495

Qy    536 gnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaayccnc 595
          | : :  :  :|:     :   :::: | ::: | : : | |      ::      :
Db    496 garggnccntgyytnwsnytngarwsngarytngargtnatggcnacnwsnmgngayaar 555

Qy    596 argayaayggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnathaarw 655
            ::  : :    :  : | :  ||        ::|  :  :  ::  ::  :|  |
Db    556 gargarytnytntgggcntggcarggntggcargaygcngtnggnmgncarathtgyacn 615

Qy    656 snytngcntggmgnccnmgnmgnaartggttytgyggnacnggnccnggnwsnytntgyt 715
            |       :  :   | :   ::    :   |: | | :      |  :   |
Db    616 acnttygarcaytaygtngarytnwsnaayaargcngcncarytnaayggngtnnnnaar 675
```

```
Qy    716 gygtncarccnmgngayytngtnccntgygtnccngtnaaywsngcngtngcnwsngarg 775
Db    676 gayatgggngcnytntggcaywsnaartaygarwsngayacnytngarcargayytngar 735

Qy    776 gngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarw 835
Db    736 mgnytnttycargarytnmgnccnytntayytnaayccncayacntaygtnmgnmgngcn 795

Qy    836 snmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntgga 895
Db    796 ytncaymgncaytayggnccngarytnathgayytnmgnggnccnathccngcncayytn 855

Qy    896 tgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtng 955
Db    856 ytnggngaraayacnytngcncarwsntgggtnaayathytngayccngtnytnccntty 915

Qy    956 tncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnw 1015
Db    916 ytnaaraarathccngargaygtnacnaarathatgaargrtncarcaytggaarccngar 975

Qy   1016 snmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayw 1075
Db    976 aarytnatgytngargargcngaracnttyttyacntayytnggnytngcnytnccnccn 1035

Qy   1076 snytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcng 1135
Db   1036 gcnccnccnwsnttytggaaraarytnatgytnatgmgnccnacngayggnmgngargtn 1095

Qy   1136 gnatg 1140
Db   1096 gartg 1100


RESULT 13
US-09-351-414-3
; Sequence 3, Application US/09351414
; Patent No. 6265199
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT: Baindur, Nand
;  APPLICANT: Deisher, Theresa A.
;  APPLICANT: Bishop, Paul D.
;. TITLE OF INVENTION: DISINTEGRIN HOMOLOG
;  FILE REFERENCE: 98-29
;  CURRENT APPLICATION NUMBER: US/09/351,414
;  CURRENT FILING DATE: 1999-07-09
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 2088
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: zdintl amino acid degenerate sequence
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (1)...(2088)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(2088)
;   OTHER INFORMATION: n = A,T,C or G
US-09-351-414-3


  Query Match             3.5%;  Score 48.8;  DB 4;  Length 2088;
  Best Local Similarity   11.8%;  Pred. No. 0.0026;
  Matches  140;  Conservative  286;  Mismatches  749;  Indels    8;  Gaps    3;

Qy    190 aarytnwsnggnaarytnccnytnccnttyaarccnathathttyacnggngtnytntay 249
Db    731 aratgytncaygarttywsnaartaymgncarmgnathaarcarcaygcngaygcngtnc 790

Qy    250 aaygcncarmgngayytnaargargcnatgggngtnttygcntgymgngtnccnggnaay 309
Db    791 ayytnathwsnmgngtnacnttycaytayaarmgnwsnwsnytnwsntayttygarggng 850

Qy    310 taytaywsnwsnttygaygtngarytncaycaytgyaargtnaayathtggytnatgmgn 369
Db    851 tntgywsnmgnacnmgnggngtnggngtnaaygartayggnytnccnatggcngtngcnc 910

Qy    370 aarcarathytngcnaayaargargarathwsnaarcarcarwsnathcargargtnacn 429
Db    911 argtnytnwsncarwsnytngcncaraayytnggnathcartgggarccnwsnwsnmgna 970

Qy    430 tgggtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnwsngaraay 489
Db    971 arccnaartgygaytgyacngarwsntggggnggntgyathatggargaracnggngtnw 1030

Qy    490 ytncayccngayaaygtnathaaraaraaraayccnttywsngarggnaarttyaarytn 549
Db   1031 sncaywsnmgnaarttywsnaartgywsnathytngartaymgngayttyytncarmgng 1090

Qy    550 gcngcngarathtgyathtgyaaygargarytnaaygtnaayccncargayaaygggngar 609
Db   1091 gnggnggngcntgyytnttyaaymgnccnacnaarytnttygarccnacngartgyggna 1150

Qy    610 aayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgn 669
Db   1151 ayggntaygtnga--rgcnggngargartgygaytgyggnttycaygtngartgytaygg 1208

Qy    670 ccnmgnmgnaartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgn 729
Db   1209 nytntgytgyaaraartgywsnytnwsnaayggngcncaytgywsng-ayggnccntgyt 1267

Qy    730 gayytngtnccntgygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaar 789
Db   1268 gyaayaayacnwsntgyytnttycarccnmgnggntaygartgymgngaygcngtnaayg 1327

Qy    790 ccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtn 849
Db   1328 artgygayathacngartaytgyacnggngaywsnggncartgyccnccnaayytncaya 1387

Qy    850 tgggarccnccnathmgnttycaraarathtayggnaaycentggatgccnmgncaraar 909
Db   1388 arcargayggntaygcntgyaaycaraaycarggnmgntgytayaayggngartgyaara 1447

Qy    910 ttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaay 969
Db   1448 cnmgngayaaycartgycartayathtggggnacnaargcngcnggnwsngayaarttyt 1507

Qy    970 gtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgn 1029
Db   1508 gytaygaraarytnaayacngarggnacngaraarggna-----aytgyggnaargaygg 1562

Qy   1030 mgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtn 1089
Db   1563 ngaymgntggathcartgywsnaarcaygaygtnttytgyggnttyytnytntgyacnaa 1622

Qy   1090 ccngaraarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtn 1149
Db   1623 yytnacnmgngcnccnmgnathggncarytncarggngarathathccnacnwsnttyta 1682

Qy   1150 ccntayaaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgy 1209
Db   1683 ycaycarggnmgngtnathgaytgywsnggngcncaygtngtnytngaygaygayacnga 1742

Qy   1210 caygayytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygay 1269
Db   1743 ygtnggntaygtngargayggnacnccntgyggnccnwsnatgatgtgyytngaymgnaa 1802

Qy   1270 tgyccnacnggnttymgnacntayatgggnccngtnccnytntgyttyggncarttytty 1329
Db   1803 rtgyytncarathcargcnytnaayatgwsnwsntgyccnytngaywsnaarggnaargt 1862

Qy   1330 ccnttyggnacngcngtnttyacncartgyytntayytncayt 1372
```

```
            ::: |  |        |     |::|     |: | :|::
Db   1863 ntgywsnggncayggngtntgywsnaaygargcnacntgyath 1905


RESULT  14
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17


  Query Match             3.5%;  Score 48.6;  DB 4;  Length 289;
  Best Local Similarity   10.1%;  Pred. No. 0.00044;
  Matches   27;  Conservative  101;  Mismatches  139;  Indels    0;  Gaps    0;

Qy    389 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnt 448
          ::: | ::::::: :|||:: ||:: : :|:::||   :   :::         :::
Db     19 rarcrurarurururarcrararururarcrararurgrnrnrsrnrnrsrnrnrsrnrn 78

Qy    449 tywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtna 508
           ::    :: : : : |   : | ::: | : :   :::           :   :::
Db     79 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 138

Qy    509 thaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyatht 568
           :: | ::: | ::     ::   :::   ::    ::       :   ::: : ::: :
Db    139 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 198

Qy    569 gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 628
          :::   ::: |   :     :       | : :  |: :|||:::: :   :    | |
Db    199 rsrnrnrsrnrnrsrnrnrsrcrargrcrurgrcrgrurararcrurcrururgrcrgrc 258

Qy    629 armgnwsnwsncarcarwsnathaarw 655
          : :  :  :   |: |::  | :||::
Db    259 ruaaaaaaaaaaaaaaaaaaaaaaaa 285


RESULT  15
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17


  Query Match             3.5%;  Score 48.6;  DB 4;  Length 289;
  Best Local Similarity   10.1%;  Pred. No. 0.00044;
  Matches   27;  Conservative  101;  Mismatches  139;  Indels    0;  Gaps    0;

Qy    389 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnt 448
          ::: | ::::::: :|||:: ||:: : :|:::||   :   :::         :::
Db     19 rarcrurarurururarcrararururarcrararurgrnrnrsrnrnrsrnrnrsrnrn 78

Qy    449 tywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtna 508
           ::    :: : : : |   : | ::: | : :   :::           :   :::
Db     79 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 138

Qy    509 thaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyatht 568
           :: | ::: | ::     ::   :::   ::    ::       :   ::: : ::: :
Db    139 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 198

Qy    569 gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 628
          :::   ::: |   :     :       | : :  |: :|||:::: :   :    | |
Db    199 rsrnrnrsrnrnrsrnrnrsrcrargrcrurgrcrgrurararcrurcrururgrcrgrc 258

Qy    629 armgnwsnwsncarcarwsnathaarw 655
          : :  :  :   |: |::  | :||::
Db    259 ruaaaaaaaaaaaaaaaaaaaaaaaa 285


Search completed: July  8, 2002, 16:20:40
Job time: 12164 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 13:28:40 ; Search time 3393.27 Seconds
(without alignments)
5477.108 Million cell updates/sec

Title: US-09-997-610-3
Perfect score: 1377
Sequence: 1 athgtngtnathccngtnyt.....gyytntayytncaytgyatg 1377

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estp1:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	479.2	34.8	2615	10	BE420422	BE420422 32-393 hu
	2	450.2	32.7	2009	11	BC004496	BC004496 Homo sapi
	3	398	28.9	1514	10	BE512633	BE512633 32-1514 h
	4	336.8	24.5	870	10	BI488505	BI488505 603021014
	5	335.6	24.4	1050	10	BM472108	BM472108 AGENCOURT
	6	321.8	23.4	1036	10	BM471183	BM471183 AGENCOURT
	7	318.6	23.1	898	10	BM457166	BM457166 AGENCOURT
	8	313	22.7	717	12	AG086951	AG086951 Pan trog1
c	9	306.6	22.3	736	10	BI754555	BI754555 603023384
	10	300.4	21.8	654	12	AG112666	AG112666 Pan trog1
	11	296	21.5	676	12	AG064424	AG064424 Pan trog1
	12	287.4	20.9	691	12	AG062447	AG062447 Pan trog1
	13	286.2	20.8	720	12	AG116938	AG116938 Pan trog1
	14	283.8	20.6	623	12	AQ037711	AQ037711 CIT-HSP-2
	15	283.2	20.6	677	12	AG097258	AG097258 Pan trog1
	16	282.4	20.5	704	12	AQ588828	AQ588828 CITBI-E1-
	17	281.6	20.5	659	12	AG092858	AG092858 Pan trog1
	18	281.4	20.4	701	12	AQ588873	AQ588873 CITBI-E1-
	19	281	20.4	729	10	BM045873	BM045873 603624875
	20	279.6	20.3	760	10	BG281182	BG281182 602402107
	21	279.4	20.3	697	12	AG018926	AG018926 Homo sapi
	22	278.8	20.2	885	10	BG761979	BG761979 602718930
	23	277.2	20.1	1182	10	BM451317	BM451317 AGENCOURT
	24	276.2	20.1	639	12	AG115932	AG115932 Pan trog1
	25	275.2	20.0	676	12	AG125093	AG125093 Pan trog1
	26	275	20.0	721	10	BE728616	BE728616 601563827
	27	274.8	20.0	671	12	AG045131	AG045131 Pan trog1
	28	274.6	19.9	877	10	BI916691	BI916691 603178655
	29	272.2	19.8	625	12	AG070563	AG070563 Pan trog1
	30	271.4	19.7	704	12	AQ020199	AQ020199 CIT-HSP-2
	31	271.2	19.7	686	12	AG119684	AG119684 Pan trog1
	32	268.8	19.5	660	12	AG057341	AG057341 Pan trog1
c	33	268.6	19.5	746	12	AG001408	AG001408 Homo sapi
	34	266	19.3	623	12	AG075663	AG075663 Pan trog1
c	35	266	19.3	699	9	AL602279	AL602279 DKFZp686B
	36	264.8	19.2	543	12	B94403	B94403 CIT-HSP-217
	37	263.8	19.2	675	12	AG096655	AG096655 Pan trog1
	38	261.4	19.0	833	9	AU140358	AU140358 AU140358
	39	261	19.0	771	12	AG001229	AG001229 Homo sapi
	40	260.6	18.9	716	12	AG104336	AG104336 Pan trog1
c	41	260.6	18.9	940	10	BG679573	BG679573 602627616
	42	260.6	18.9	955	10	BG681769	BG681769 602627877
	43	257.4	18.7	649	9	AW857140	AW857140 RC1-CT030
	44	256.6	18.6	900	10	BM449699	BM449699 AGENCOURT
	45	255.4	18.5	676	12	AQ239408	AQ239408 CIT-HSP-2

ALIGNMENTS

RESULT 1
BE420422
LOCUS BE420422 2615 bp mRNA linear EST 11-OCT-2001
DEFINITION 32-393 human bone marrow cDNA library Homo sapiens cDNA, mRNA sequence.
ACCESSION BE420422
VERSION BE420422.1 GI:16041640
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2615)
AUTHORS Lu,X., Cui,L. and Li,Y.
TITLE DDRT-PCR from B cell
JOURNAL Unpublished (2000)
COMMENT Contact: xingwu Lu,liangxian Cui,yonghai Li
Department of Biochemistry
Institute of Basic Medical Science, Peking Union Medical College
DongDan SanTiao 5, Beijing, P.R.C, 100005
Tel: 86-010-65296951
Email: luxingwu@263.net
full-length and coding sequence.
FEATURES Location/Qualifiers
source 1. .2615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="human bone marrow cDNA library"
/tissue_type="bone marrow"
BASE COUNT 705 a 531 c 641 g 738 t
ORIGIN

Query Match 34.8%; Score 479.2; DB 10; Length 2615;
Best Local Similarity 47.1%; Pred. No. 4.6e-98;
Matches 500; Conservative 222; Mismatches 299; Indels 41; Gaps 4;

Qy 352 aayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcar 411
Db 494 AACCTTGAGCTTGAGAGAGATCATTTAGGGTATCTGGCAGAAGAAATTTCTAAGCAGCAA 553

```
Qy    412 wsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngar 471
          :: ||:||:||:|| || ||||| :| :| ||:|| ||::: ||:||::| ||:|| | :
Db    554 AGCATTCAAGAGGTAACTTGGGTGCTGTTAAAGGCATTTAGTTTTATAAGGGAAGCAGGG 613

Qy    472 cayaarwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccnttywsn 531
          ||: |::: :  ||:||::| || || ||:||:|  ||: |:||:||:| :|| | :::
Db    614 CATGAAAGTTTGGAAAATTTGCAGCCTGACAATGCAATAGAAAAGAAAATCCCATATTCT 673

Qy    532 garggnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaay 591
          ||:|  ||: |:||::| || || ||:||:||:||:||:||: |:||::| ||:|| ||:
Db    674 GAGGAGAAAATCAAGCTGGCTGCAGAAATTTGCATATGTAACAAGGAGTTAAATGTTAAT 733

Qy    592 ccncargayaaygggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnath 651
          || ||:||:||:|| ||:||: |:::  || | ||:||::| :: :: ||:||: :  |:
Db    734 CCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGCCAGAGGTCTTCACAGCAGCCCCTC 793

Qy    652 aarwsnytngcntggmgnccnmgnmg---naartggttytgyggnacnggnccnggnwsn 708
            ::: :  || || :| || :| :|    ||:|| | :  :||  | || || || ::
Db    794 CCATCACAGGCCTGCAGGCCTAGGAGGAAAAAGTGATCTCTTGGGCCAGGCCCAGGGTCC 853

Qy    709 ytntgytgygtncarccnmgngayytngtnccntgygtnccngtnaaywsngc------- 761
          :  ||:||:|| ||:||  | ||::| || || ||:|| || |    : : ||
Db    854 CCATGCTGTGTGCAGCCTGGGGACTTGGTGCCTTGTGTCCCAGCTGCTCCAGCCATGGCT 913

Qy    762 -------------------------ngtngcnwsngarggngcnwsnccnaarccntgg 795
                                    || || :: ||:|| || ::    ||:|| |||
Db    914 GAAAGGGGCCAACATAGAGCTCGGATTGTGGCTTCAGAGGGTGCAAGCTGTAAGCCTTGG 973

Qy    796 carytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggar 855
          ||::| || :: || || ||:|| |   | ||  |:||::: :  ||:| :|| |||||:
Db    974 CAGCTTCCATGTGGTGTTGAGCCTGCCAGTGCACAGAAGTCAAAAATTGGGGTTTGGGAA 1033

Qy    856 ccnccnathmgnttycaraarathtayggnaaycentggatgccnmgncaraarttygcn 915
          || |    ::| ||:||:||:|| ||:|| ||: | ||||||||  | ||:||:| :||
Db   1034 CCTCTGCCTAGATTTCAGAAGATGTATGGAAATGCCTGGATGCCCTGGCAGAAGTCTGCT 1093

Qy    916 gtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggn 975
          || || |  ||  : :: |||:| |  :: || :| |  || ||:||:|| ||:|| ||
Db   1094 GTAGGGGAGGGGCCCTCATGGAGAATCTCTGCTAGGGCAGTACAGAAGGGAAATGTGGGA 1153

Qy    976 tgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsn 1035
          | |||:|| |  ||::|  | || :: || || ||  : ::  | || || :| :|  :
Db   1154 TCGGAGCCCCAGCACAGAATCCCTACTGGGGCACCGCCTAGTGGAGCTGTGAGAAGAGGG 1213

Qy   1036 ccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngar 1095
          || || :: :: :| :  ||:|| |  :| :: || ||::: :| |  ||:|  || | :
Db   1214 CCACCATCCTCCAGACCCCAGAATGATAGATCCACTGACAGCTTGCTCCATGCACCTGGA 1273

Qy   1096 aarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntay 1155
          ||:::  | ||    ||:  :||:|| | ||:|| ||  |   |||: : || || | :
Db   1274 AAAACCGCAGA----CAACACCAGCCCATGAAAGCAGCTAGGCAGGAGGCTGTACCCTGC 1329

Qy   1156 aaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygay 1215
          ||: |    |  || ||::|  | ||:|   | || | : |::| :| ||:   ||:||:
Db   1330 AAAGCCACAGGGGCGGAGCTGCCCAAGATCATGGGAACCCACCTCTTGCATCAGCATGAC 1389

Qy   1216 ytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccn 1275
          :| ||:|| :|   :|| || ||: | ||:||:||:|| || :| :| ||:||:||:||
Db   1390 CTGGATGTGAGATTTGGAGTTAAAGGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCC 1449

Qy   1276 acnggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccntt 1334
            |  | ||::  || | :|||||| || ||  || :| ||:||:  ||:||: |:|| ||
Db   1450 TCTAGATTTCAGACTTGCATGGGGCCTGTAGCCCCTTTGTTTTACTCAATTTCTCCCATT 1509

Qy   1335 yggnacngcngtnttyacncartgyytntayytncaytgyat 1376
          :|| || || || ||:|| ||:| ::| ||::  ||:||:||
Db   1510 TGGAACAGCTGTGTTTACACAATACCTGTACCCCCATTGTAT 1551


RESULT  2
BC004496
LOCUS       BC004496                2009 bp    mRNA    linear   HTC 12-JUL-2001
DEFINITION  Homo sapiens, Similar to hypothetical protein FLJ14058, clone
            IMAGE:3831313, mRNA.
ACCESSION   BC004496
VERSION     BC004496.1  GI:14709139
KEYWORDS    HTC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2009)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 14 Row: d Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1. .2009
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3831313"
                     /tissue_type="Skin, melanotic melanoma."
                     /clone_lib="NIH_MGC_20"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
BASE COUNT      556 a    429 c    513 g    511 t
ORIGIN


  Query Match             32.7%;  Score 450.2;  DB 11;  Length 2009;
  Best Local Similarity   45.2%;  Pred. No. 1.6e-91;
  Matches  452;  Conservative  214;  Mismatches  321;  Indels   13;  Gaps    3;

Qy    390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
          :|||:||:||::: | :|||:||::: ||:||:||: | || |||| :| :| ||:|| ||
Db    489 AGAAGAAATTTCTATGCAGCAAAGCATTCAAGAGTTGACTTGGGTGCTGTTAAAGGCATT 548

Qy    450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
          ::: ||:||::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||
Db    549 CAGATTTATAAGGGAAGCAGAGCATAAAAGTTCAGAAAATTTGCAGCCTGACAATGTGAT 608

Qy    510 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 569
          : |:||:||:| :|| ||::  ||:|  ||:||:||::  || ||  |:||:||:||: |
Db    609 AGAGAAGAAAAGCCCATTTTTTGAGGAGAAATTCAAGCCAGCTGCATAAATTTGCATAAG 668

Qy    570 yaaygargarytnaaygtnaayccncargayaaygggngaraayathwsntggacntgyca 629
          :||: |: |::   |:|| ||:|| ||:||:||:|| ||: |:||::: ||| | :||
Db    669 TAACAAGAAGCCGCATGTTAATCCCCAAGACAACGGAGAAGACATCTCCAGGACATATCA 728

Qy    630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgngaartggttytg 689
          ::| :: ::'||:||: :  |:  ::: :  ||   |:| || :| :| ||:||||||:
Db    729 GAGGTCTTCACAGCAGCCCCTCCCGTCACAGGCCCAGAGGCCTAGGAGAAAATGGTTTCC 788

Qy    690 yggnacnggnccngngnwsnytnt------gytgygtncarccnmgngayytngtncc--- 740
```

```
Db    789 TGGACCTGTGCAGCCTAGGTACTTGGCACCCTGTGTCCCAGCTACTCCAGTAGTGGCTGA 848

Qy    741 ---ntgygtnccngtnaaywsngcngtngcnwsngargggngcnwsnccnaarccntggca 797
Db    849 AAGGGGCCAAAGTACAGCTTACGCCATGGCTTCAGAGGGTGCAAGCGCCAAGCCTTGGCA 908

Qy    798 rytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarcc 857
Db    909 GCTTCCGCATGCTGTTGAGCCTGCAAATGCACAGACGTCAAGAATTTGGATTTGGGAACC 968

Qy    858 nccnathmgnttycaraarathtayggnaaycentggatgccnmgncaraarttygcngt 917
Db    969 TCTGCCTAGATTTCAGGAGATGTACGGAAGTGTCCGGATGCCCAGGCAAAAGTTTGCTGC 1028

Qy    918 nggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntg 977
Db   1029 AGGGGCAGGGCCTTCATGGAGAACCTCTGCTAGGGCAGTGTGGAAGGGAAACATGGGGTC 1088

Qy    978 ggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsncc 1037
Db   1089 TGAGCCCCCACACAGAATCCCTACTGAGGCACCACCTAGTGGAGCTGTGAGAAAAGAGCC 1148

Qy   1038 nccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraa 1097
Db   1149 ACCATCCTCCACACCCCAGAATGATAGATCCACCAACAGCTTGCACCATGCACCTGGAAA 1208

Qy   1098 rwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntayaa 1157
Db   1209 AGCCGCAGACATTCAACACCAGCCCGTGAAAGCAGCCAAAAGGGGGAATATATCCTGCAA 1268

Qy   1158 racngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygayyt 1217
Db   1269 AGCCACAGGGGAGGAGCTTCCCAAGACTATGGGAACCTACCTCTTACATCAGCACGACCT 1328

Qy   1218 ngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnac 1277
Db   1329 TTGTGTGAGACCTGGAGTCAAAGGAGATCATTTTGGAGCTTTAAAATCTGACTGCCCTGC 1388

Qy   1278 nggnttymgnacntayatgggncc-ngtnccnytntgyttyggncarttyttyccnttyg 1336
Db   1389 TGGATTTCAGACTTGCATGGGCCCTACAACCCCTTTGTTTTGGCCGATTTCTCCAATTTG 1448

Qy   1337 gnacngcngtnttyacncartgyytntayytncaytgyat 1376
Db   1449 GAAAGGCTGTATTTACCCAATACCTGTACCCCCATTGTAT 1488
```

```
RESULT   3
BE512633
LOCUS       BE512633               1514 bp    mRNA    linear   EST 11-OCT-2001
DEFINITION  32-1514 human B lymphocyte cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   BE512633
VERSION     BE512633.1  GI:16041645
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Lu,X., Cui,L. and Li,Y.
  TITLE     DDRT-PCR from B cell
  JOURNAL   Unpublished (2000)
COMMENT     Contact: xingwu Lu,liangxian Cui,yonghai Li
            Department of Biochemistry
            Institute of Basic Medical Science, Peking Union Medical College
            DongDan SanTiao 5, Beijing, P.R.C, 100005
            Tel: 86-010-65296951
            Email: luxingwu@263.net
            full cDNA sequence.
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="human B lymphocyte cDNA library"
                     /tissue_type="bone marrow"
BASE COUNT      380 a    341 c    398 g    395 t
ORIGIN
```

Query Match 28.9%; Score 398; DB 10; Length 1514;
Best Local Similarity 47.1%; Pred. No. 1.1e-79;
Matches 427; Conservative 182; Mismatches 256; Indels 41; Gaps 4;

```
Qy    508 athaaraaraaraaycentywsngarggnaarttyaarytngcngcngarathtgyath 567
Db      1 ATAGAAAAGAAAATCCCATATTCTGAGGAGAAAATCAAGCTGGCTGCAGAAATTTGCATA 60

Qy    568 tgyaaygargarytnaaygtnaayccncargayaaygggngaraayathwsntggacntgy 627
Db     61 TGTAACAAGGAGTTAAATGTTAATCCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGC 120

Qy    628 carmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naartgg 684
Db    121 CAGAGGTCTTCACAGCAGCCCCTCCCATCACAGGCCTGCAGGCCTAGGAGGAAAAAGTGA 180

Qy    685 ttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgy 744
Db    181 TCTCTTGGGCCAGGCCCAGGGTCCCCATGCTGTGTGCAGCCTGGGGACTTGGTGCCTTGT 240

Qy    745 gtnccngtnaaywsngc----------------------------------ngtngcnwsn 771
Db    241 GTCCCAGCTGCTCCAGCCATGGCTGAAAGGGGCCAACATAGAGCTCGGATTGTGGCTTCA 300

Qy    772 garggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaar 831
Db    301 GAGGGTGCAAGCTGTAAGCCTTGGCAGCTTCCATGTGGTGTTGAGCCTGCCAGTGCACAG 360

Qy    832 aarwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccn 891
Db    361 AAGTCAAAAATTGGGGTTTGGGAACCTCTGCCTAGATTTCAGAAGATGTATGGAAATGCC 420

Qy    892 tggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgn 951
Db    421 TGGATGCCCTGGCAGAAGTCTGCTGTAGGGGAGGGGCCCTCATGGAGAATCTCTGCTAGG 480

Qy    952 gtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccn 1011
Db    481 GCAGTACAGAAGGGAAATGTGGGATCGGAGCCCCAGCACAGAATCCCTACTGGGGCACCG 540

Qy   1012 wsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacn 1071
Db    541 CCTAGTGGAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAATGATAGATCCACT 600

Qy   1072 gaywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcn 1131
Db    601 GACAGCTTGCTCCATGCACCTGGAAAAACCGCAGA----CAACACCAGCCCATGAAAGCA 656

Qy   1132 gcnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggn 1191
Db    657 GCTAGGCAGGAGGCTGTACCCTGCAAAGCCACAGGGGCGGAGCTGCCCAAGATCATGGGA 716

Qy   1192 athtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycaytty 1251
Db    717 ACCCACCTCTTGCATCAGCATGACCTGGATGTGAGATTTGGAGTTAAAGGAGATCATTTT 776

Qy   1252 ggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnyt 1310
Db    777 GGAGCTTTAAGATTTGACTGCCCCTCTAGATTTCAGACTTGCATGGGGCCTGTAGCCCCT 836

Qy   1311 ntgyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytnca 1370
Db    837 TTGTTTTACTCAATTTCTCCCATTTGGAACAGCTGTGTTTACACAATACCTGTACCCCCA 896

Qy   1371 ytgyat 1376
```

```
          :||:||
Db    897 TTGTAT 902


RESULT   4
BI488505
LOCUS       BI488505                870 bp    mRNA    linear   EST 28-AUG-2001
DEFINITION  603021014F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5191803 5',
            mRNA sequence.
ACCESSION   BI488505
VERSION     BI488505.1  GI:15327733
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 870)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11479  row: k  column: 04
            High quality sequence stop: 849.
FEATURES             Location/Qualifiers
     source          1. .870
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5191803"
                     /clone_lib="NIH_MGC_114"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is  destroyed
                     upon cloning).  Average insert size 1.5 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length  clones and was constructed by C. Gruber
                     (Invitrogen).  Research  Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT      222 a    211 c    256 g    181 t
ORIGIN


  Query Match             24.5%;  Score 336.8;  DB 10;  Length 870;
  Best Local Similarity   44.6%;  Pred. No. 7.2e-66;
  Matches  378;  Conservative  161;  Mismatches  273;  Indels   35;  Gaps    3;

Qy    493 cayccngayaaygtnathaaraaraaraayccnttywsngarggnaarttyaarytngcn 552
          || || ||: |:|   |: |:||:||:||:|| ||::  ||:|  ||:|||:||::  ||
Db      2 CAGCCTGATGATGAAGTAGAAAAGAAAAACCCATTTTTTGAGGATAAATTCAAGCCAGCT 61

Qy    553 gcngarathtgyathtgyaaygargarytnaaygtnaayccncargayaayggngaraay 612
          || ||:||:||:||:||:| : |:||::  ||:|| ||:|  ||:||: |:|| ||:||:
Db     62 GCAGAAATTTGCATATGTAGCAAGGAGTCTAATGTTAATCTCCAAGACCATGGGGAAAAT 121

Qy    613 athwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnc-c 671
          |:::  || | ||:||::|  : ::   :||: :  |:  ::: :  ||  |  | |
Db    122 GTCTCCGGGCCATGTCAGAGACCTTCATGGCAGCCCCTCCCATCACAGGCTGGAGGCCAG 181

Qy    672 nmgnmgnaartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgnga 731
          :| :  ||: |||| |:||  | || || |  :: :  ||:| :|| ||:|| :| ||
Db    182 GAGGAAAAAGAGGTTTCGTGGGCCAGGCCCAGCATCCCCGTGCTTTGTGCAGCCTAGGGA 241

Qy    732 yytngtnccntgygtnccngt----------------------------------naayws 758
          ::| || || ||:|| || ||                                    : :

Db    242 CTTGGTGCCCTGTGTCCCAGTCGCTCCAGCTGTGACTGAAAGGGGCCAACGTAGAGCTCG 301

Qy    759 ngcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarcc 818
          || || || :: ||:|| |  ::  | ||:|  |||||::| ||  : || || ||:||
Db    302 GGCTGTGGCTTCAGAGGGTGGAAGCTCTAAGCTTTGGCAGCTTCCACGTGGTGTTGAGCC 361

Qy    819 ngtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarat 878
            || ||  |:||::: :| | :||:|| |||||:|| |    ::| ||:||:||:||
Db    362 TACAGGTGCACAGAAGTCAAGAACTGAGGTTTGGGAACCTCTGCCTAGATTTCAGAAGAT 421

Qy    879 htayggnaaycntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmg 938
          | :|| ||: | |||||||| :| ||:||:||:|| |  || || ||  : :  |||:|
Db    422 GTTTGGAAATGCCTGGATGCCCAGGCAAAAGTTTGCTGCAGGAGTGGGGCCCTTATGGAG 481

Qy    939 nacnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtncc 998
          || :: ||  | |  || ||:||:|  ||:|| || | |||:|| |  || :| || |
Db    482 AACCTCTGCTGGAGCAGTGCAGAAGGAAAATGTGGGATCGGAGCCCCGACAGAGAGTCCA 541

Qy    999 nwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraa 1058
          :: || || |   : ::  |  | || :| :|    || || :: :: :| :  ||:||
Db    542 TACTGGGGCACTGCCTAGTGGAACTGTGAGAAGACAGCCACCGTCCTCCAGACCCCAGAA 601

Qy   1059 rggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgyca 1118
          || :| :: ||  |::: :| || ||:|| || | :||:    | | :|| ||:||:||
Db    602 TGGTAGATCCACCAACAGCTTGCACCATGTGCCTGGAAAAGATGCAGGTACTCAATGCCA 661

Qy   1119 rccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnac 1178
          :||  | ||:|| ||  | | |||: : || || | :||: |    |     ||::|  |
Db    662 GCCCATGAAGGCAGCCAGGAGGGAGGCTGTACCCTGCAAAGCCACAGGGAGGGAGCTGTC 721

Qy   1179 naaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncaygggngtnaa 1238
          ||:||  | || | : |::  :    : |:     ::| ||:|  :| | :|| || ||
Db    722 CAAGACCATGGGAACCCACTCTTGCATCAGCGTGACCTTGGATGCGAGACCTGGAGTCAA 781

Qy   1239 rmgngaycayttyggngcnyt-nmgnttygaytgyccnacnggnttymgnacntayatgg 1297
          : | ||:||:||:|| || :|  :  | :||:   ||  | || ||::  || | :|| |
Db    782 AGGAGATCATTTTGGAGCTTTAACAATGTGACCTGCCTGCTGGATTTCAGACTTGCATTG 841

Qy   1298 gnccngt 1304
          |  |  |
Db    842 GGGCCCT 848


RESULT   5
BM472108
LOCUS       BM472108                1050 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6465359 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5539381
            5', mRNA sequence.
ACCESSION   BM472108
VERSION     BM472108.1  GI:18521150
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1050)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12233  row: m  column: 14
            High quality sequence stop: 658.
FEATURES             Location/Qualifiers
     source          1. .1050
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5539381"
                     /clone_lib="NIH_MGC_72"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      295 a    262 c    272 g    221 t
ORIGIN


  Query Match             24.4%;  Score 335.6;  DB 10;  Length 1050;
  Best Local Similarity   43.8%;  Pred. No. 1.5e-65;
  Matches  335;  Conservative  169;  Mismatches  248;  Indels   12;  Gaps    2;

Qy    390 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 449
          :||:||:||::: | :||:||::: ||:||:||: | || ||||| :| :| ||:|| ||
Db     16 AGAAGAAATTTCTATGCAGCAAAGCATTCAAGAGTTGACTTGGGTGCTGTTAAAGGCATT 75

Qy    450 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 509
          ::: ||:||::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||
Db     76 CAGATTTATAAGGGAAGCAGAGCATAAAAGTTCAGAAAATTTGCAGCCTGACAATGTGAT 135

Qy    510 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 569
          : |:||:||:| :|| ||::  ||:|  ||:||:||::  || ||  |:||:||:||: |
Db    136 AGAGAAGAAAAGCCCATTTTTTGAGGAGAAATTCAAGCCAGCTGCATAAATTTGCATAAG 195

Qy    570 yaaygargarytnaaygtnaayccncargayaaygggngaraayathwsntggacntgyca 629
          :||: |: |::   |:|| ||:|| ||:||:||:|| ||: |:||:::  |||| | :||
Db    196 TAACAAGAAGCCGCATGTTAATCCCCAAGACAACGGAGAAGACATCTCCAGGACATATCA 255

Qy    630 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 689
          ::| :: :: ||:||: :  |:  ::: :  ||   |:| || :| :| ||:|||||:
Db    256 GAGGTCTTCACAGCAGCCCCTCCCGTCACAGGCCCAGAGGCCTAGGAGAAAATGGTTTCC 315

Qy    690 yggnacnggnccnggnwsnytnt------gytgygtncarccnmgngayytngtncc--- 740
          :||  | |  |     :: :  |       :||:|| | : | :       | ||  |
Db    316 TGGACCTGTGCAGCCTAGGTACTTGGCACCCTGTGTCCCAGCTACTCCAGTAGTGGCTGA 375

Qy    741 ---ntgygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggca 797
             |:          ::  ||  | || :: ||:|| || ::  | ||:|| |||||
Db    376 AAGGGGCCAAAGTACAGCTTACGCCATGGCTTCAGAGGGTGCAAGCGCCAAGCCTTGGCA 435

Qy    798 rytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarcc 857
          ::| ||    |  || ||:|| |     ||  |:| ::: :| ||:  : ( |||||:||
Db    436 GCTTCCGCATGCTGTTGAGCCTGCAAATGCACAGACGTCAAGAATTTGGATTTGGGAACC 495

Qy    858 nccnathmgnttycaraarathtayggnaaycentggatgccnmgncaraarttygcngt 917
            |    ::| ||:||: |:|| ||:|| | :    ||||||| :| ||:||:||:|| |
Db    496 TCTGCCTAGATTTCAGGAGATGTACGGAAGTGTCCGGATGCCCAGGCAAAAGTTTGCTGC 555

Qy    918 nggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntg 977
            || |  ||  : :: |||:| || :: ||  :|  || ||  : ::  | || ||:| :     ||
Db    556 AGGGGCAGGGCCTTCATGGAGAACCTCTGCTAGGGCAGTGTGGAAGGGAAACATGGGGTC 615

Qy    978 ggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsncc 1037
            ||:|| || ||::|  | || :: |  || ||  : ::  | || || :| :     ||
Db    616 TGAGCCCCCACACAGAATCCCTACTGAGGCACCACCTAGTGGAGCTGTGAGAAAAGAGCC 675

Qy   1038 nccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraa 1097
            || :: :: :  :  ||:|| |  :| :: ||  |::: :| || ||:|  ||   :||
Db    676 ACCATCCTCCACACCCCAGAATGATAGATCCACCAACAGCTTGCACCATGCACCCTGGAA 735

Qy   1098 rwsnacngayacncartgycarccngtnaargcngcnggnatgg 1141
          : :  | ||:|  ||:  :||:|| |   |:|| ||    | ||
Db    736 AGCCGCAGACATTCAACACCAGCCCGGGGAAGCAGCCAAAAGGG 779


RESULT  6
BM471183
LOCUS       BM471183               1036 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6478314 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5563049
            5', mRNA sequence.
ACCESSION   BM471183
VERSION     BM471183.1  GI:18520225
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1036)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12293  row: g  column: 18
            High quality sequence stop: 661.
FEATURES             Location/Qualifiers
     source          1. .1036
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5563049"
                     /clone_lib="NIH_MGC_72"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      254 a    239 c    294 g    247 t      2 others
ORIGIN


  Query Match             23.4%;  Score 321.8;  DB 10;  Length 1036;
  Best Local Similarity   45.0%;  Pred. No. 2.1e-62;
  Matches  390;  Conservative  166;  Mismatches  267;  Indels   43;  Gaps    6;

Qy    553 gcngarathtgyathtgyaaygargaryt--naaygtnaayccncargayaayggngara 610
          || ||:| :| :||: |:| : |:||::|   ||:|| ||:|| ||::||: |:|| ||:|
Db      1 GCAGAAAATTACATAAGTAGCAAGGAACTACTAATGTTAATCCCCAAGACCATGGGGAAA 60

Qy    611 ayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnc 670
          |: |:::  |  | ||:||::|  : :: | :||:::  |:  ::: :  ||   |:| |
Db     61 ATGTCTCCAGTTCCTGTCAGAGACCTTCACTGCAGACCCTCCCATCACAGGCCCAGAGGC 120

Qy    671 cnmgnmg---naartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnm 727
          | :| :|    ||:|||||:||:||    ||  | || :: :  ||:||:||  |:|| :
Db    121 CCAGGAGGAAAAAGTGGTTCTGTGGGCTGGGTTCAGGGTCCCCATGCTGTGTGAAGCCTA 180

Oy    728 gngayytngtnccntgygtnccng----------------------------------tna 754
          | ||::| || || ||:|| || |                                   |
Db    181 GGGACTTGGTGCCCTGTGTCCCAGCTGCTCCAGCCTTGGCTGAAAGGGGCTAATGTATAG 240

Qy    755 aywsngcngtngcnwsngarggngcnwsnccnaarccntggcar---ytnccnwsnggng 811
            :   || || || :  ||:|| |  :: || ||:|| |||||:   :| || :: || |
Db    241 CTCTGGCTGTGGCTTTAGAGGGTGGAAGCCCCAAGCCTTGGCAGCTTCTTCCATGTGGTG 300

Oy    812 tngarccngtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttyc 871
          | ||:|| || || ||  |:||::: :| | :||:|| |||||:|| |   | :| ||:|
Db    301 TTGAGCCTGTGGGTGCACAGAAGTCAAGAACTGAGGTTTGGGAACCTCTGCT-AGATTCC 359

Qy    872 araarathtayggnaaycentggatgccnmgncaraarttygcngtnggngtnggnwsnw 931
          |:||:|| ||:|| ||: | |||||||| :| ||:||:||:|| |   |  | ||  : :
Db    360 AGAAGATGTATGGAAATGCCTGGATGCCCAGGCAAAAGTTTGCTGCAAGGTTGGGGCCCT 419

Oy    932 sntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaym 991
```

```
          : |||:| || :: || :| |  ||   :||:|| ||:|| || | |||:|| || ||::
Db    420 CATGGAGAACCTCTGCTAGGGCAGTGTGGAAGGGAAATGTGGGGTTGGAGCCCCCGCACA 479

Qy    992 gngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgny 1051
          | |  ||  : || || |   : ::  | || || :| :|  : || || :: :: :| :
Db    480 GTGCCCCTGCTGGGGCACTGCCTAGTGGAGCTGTGAGAAGAGGGCCACCATCCTCCAGAC 539

Qy   1052 tncaraargggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacnc 1111
            ||:|| || :| :: |   |::: :| ||   :|  |  | :||: :  | ||:|| |
Db    540 CCCAGAATGGTAGATCCATCAACAGCTTGCACTGTGCACTTGGAAAAGCCGCAGACACTC 599

Qy   1112 artgycarccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcng 1171
           :  :||:||  | ||:|| || || | |||: :  | || | :||:       |  || |
Db    600 GACACCAGCCCATGAAAGCAGCTGGGAGGGAGGCTATACCCTGCAAAGACACAGGGGCAG 659

Qy   1172 arytnacnaaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayg 1231
          |::|  | ||:||  | || | : |::| :| ||:   ||:||::| ||: | :| | :|
Db    660 AGCTGCCTAAGACCATGGGAACCCACCTCTTGCATCAGCATGACCTGGATATGAGACTTG 719

Qy   1232 gngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacnt 1291
          | || ||: | ||:||:||:|| || :| :  ||:||:||::||  | || ||::  || |
Db    720 GAGTCAAAGGAGATCATTTTGGAGCTTTGAAATTTGATTGCCCCGCTGGATTTCAGACTT 779

Qy   1292 ayatgggnccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtntty 1350
           : |  | |    |  || :| ||:||:|| ||:||: |:|| ||: | |  ||    ||:
Db    780 GTGTANGCCTGATAACCCCTTTGTTTTGGGCAATTTCTCCCATTTTGAATGGCTACATTT 839

Qy   1351 acncartgyytntayytncaytgyat 1376
             ||:| ::| ||::   |: | ||
Db    840 TACCAATACCTGTACCCCTATGGGAT 865


RESULT   7
BM457166
LOCUS       BM457166                 898 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6411690 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583427
            5', mRNA sequence.
ACCESSION   BM457166
VERSION     BM457166.1  GI:18506206
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12346  row: h  column: 20
            High quality sequence start: 9
            High quality sequence stop: 713.
FEATURES             Location/Qualifiers
     source          1. .898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5583427"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      225 a    228 c    254 g    191 t
ORIGIN


  Query Match             23.1%;  Score 318.6;  DB 10;  Length 898;
  Best Local Similarity   45.1%;  Pred. No. 1e-61;
  Matches  389;  Conservative  162;  Mismatches  253;  Indels   59;  Gaps    5;

Qy    562 tgyathtgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntgg 621
          ||:||: |:| : |:||::  ||:|| ||:|| ||:||: |:|| ||:||: |:::  ||
Db     35 TGCATAAGCAGCAAGGAGCCTAATGTTAATCCCCAAGACCATGGGGAAAATGTCTCCAGG 94

Qy    622 acntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgncccnmgnmg---n 678
           | ||:||::|  : ::   :||: :  |:  ::: :  ||   |:| || :| :|
Db     95 TCATGTCAGAGACCTTCATGGCAGCCCCTCCCATCACAGGCCCAGAGGCCCAGGAGGAAA 154

Qy    679 aartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtn 738
          ||:|||||:||:||    || || || :: :| ||:||:|| ||:|| :| ||::| ||
Db    155 AAGTGGTTTTGTGGGCTGGGCCCAGGGTCTCTGTGCTGTGTGCAGCCTAGGGACTTGGTG 214

Qy    739 ccntgygtnccng----------------------------------tnaaywsngcngtn 765
          || ||:|| || |                                     ::: || ||
Db    215 CCCTGTGTCCCAGCTGCTCCAGCCATGGCTGAAAGGAGCCAACATAGAGCTTGGGCTGTG 274

Qy    766 gcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggn 825
          || :: ||:||     : || ||:|  |||||::| || :: |  || ||:|| || ||
Db    275 GCTTCAGAGGGTAGAGGCCCCAAGCATTGGCAGCTTCCATGTGATGTTGAGCCTGTGGGT 334

Qy    826 gcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggn 885
          ||  |:| ::: :| ||:||: | |||||:|| |    ::| ||:||:|| ||:||:||
Db    335 GCAGAGATGTCAAGAATTGAGATTTGGGAACCTCTGCCTAGATTTCAGAATATATATGGA 394

Qy    886 aayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsn 945
          ||: | |||||| | :| ||:||:||:|| |  || |   | : :: ||| | || ::
Db    395 AACACCTGGATGTCCAGGCAAAAGTTTGCTGCAGGGGCAGAGCCCTCATGGGGAACCTCT 454

Qy    946 gcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggn 1005
           | :| |  ||   :||:|| ||:|| ||    ||:|| || ||::| || ||    ||
Db    455 CCTAGGGCAGTGTGGAAGGGAAATGTGGGG-CAGAGCCCCCACACAGAGTCCCTGTTGGG 513

Qy   1006 gcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnws--------------------n 1044
          || |   : ::  | || || :| :| :: || |  ::
Db    514 GCACTGCCTAGTGGAGCTGTGAGAAGAAGGCCACAGTCCCAGAACGGTAGACCCCAGAAT 573

Qy   1045 wsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacn 1104
           : :| :  ||:|| || :| :: ||   ::  :| || | :|| || | :||: : ||
Db    574 GGTAGACCCCAGAATGGTAGATCAACCACCAACTTGCATCTTGTGCCTGGAAAAGCTACA 633

Qy   1105 gayacncartgycarccngtnaargcngcnggnatggarwsngtnccntayaaracngtn 1164
          ||:|| ||:  :||:|| |  ||:|| ||  | | ||| : || |  | :||: |
Db    634 GACACTCAACATCAGCCTGAGAAAGCAGCCAGGAGGGAGGCTGTACTCTGCAAAGCCACA 693

Qy   1165 gtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygayytngaygtn 1224
          |  || ||::|    ||:||  | || | :||::| :| ||:   ||:||::|  |:||
Db    694 GGGGCGGAGCTGCACAAGACCATGGGAACCTATCTCTTGCATCAGCATGACCTGAATGTG 753

Qy   1225 mgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacnggntty 1284
          :| ||:|| || ||: | ||:||:| :|| || :| :  ||:||:||:|| || || ||:
Db    754 AGACATGGAGTCAAAGGAGATCATTATGGAGCTTTAAAATTTGACTGCCCCACTGGATTT 813

Qy   1285 mgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnttyggnacngc 1343
          :  || | :|||  || ||  || :| ||:||:|| ||:||: |: | ||:|| || ||
Db    814 CAGACTTGCATGGACCCTGTAACCCCTTTGTTTTGGCCAATTTCTCTCCTTTGGAACAGC 873

Qy   1344 ngtnttyacncartgyytntayy 1366
            || ||:|| ||:| ::| ||::
Db    874 TGTATTTACCCAATACCTGTACC 896


RESULT   8
AG086951
LOCUS       AG086951                 717 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-085H08.F, genomic survey sequence.
```

```
ACCESSION   AG086951
VERSION     AG086951.1  GI:16638753
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-085H08.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 717)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .717
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-085H08.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      175 a    189 c    195 g    158 t
ORIGIN
```

```
  Query Match             22.7%;  Score 313;  DB 12;  Length 717;
  Best Local Similarity   49.1%;  Pred. No. 1.7e-60;
  Matches  299;  Conservative  113;  Mismatches  195;  Indels    2;  Gaps    1;

Qy    744 ygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggcarytncc 803
          :|| |  |    :   || || || :: ||:|| || :: || ||:|| ||||||:| ||
Db     89 CGTCCGCGGAGCTCAGGCTGTGGCTTCAGAGGGTGCAAGCCCTAAGCCTTGGCAGCTTCC 148

Qy    804 nwsnggngtngarccngtnggngcnaaraaarwsnmgnathgargtntgggarccnccnat 863
            || ||  |:|| ||  | ||  |:||::: :| | :| :|| ||||||:|| |
Db    149 ACATGGTGTTAAGCCTGTGAGTGCACAGAAATCAAGAACTGGGGTTTGGGAACCTCTGCC 208

Qy    864 hmgnttycaraarathtayggnaayccntggatgccnmgncaraarttygcngtnggngt 923
          ::| ||:||:| :||:||:|| ||: | |||||||| :| ||:|| ||:|| || || |
Db    209 TAGATTTCAGAGGATATATGGTAATACCTGGATGCCCAGACAGAATTTTGCTGTAGGGGA 268

Qy    924 nggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntgggarcc 983
            ||  : :: || :  || :: || :| |  ||   : |: | ||:|| || |  ||:||
Db    269 GGGGGCCTCATGAAAAACCTCTGCTAGGGCAGTGTGGGAGAGAAATGTGGGGTCAGAGCC 328

Qy    984 nccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnws 1043
            | ||::| || || :: || || ||  : ::  | || || :  :|  : || |  :
Db    329 ATCACACAGAGTCCCTACTGGGGCACCGCCTAGTGGAGCTGTGAAAAGTGGGCCACTATT 388

Qy   1044 nwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnac 1103
            :: :| :  ||:|| || :| :: || ||::: :| || ||:|  || | :|||: : ||
Db    389 CTCCAGACCCCAGAATGGTAGATCCACTGACAGCTTGCACCATGCACCTGGAAAAGCCAC 448

Qy   1104 ngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntayaaracngt 1163
            ||:|| ||:||:||:|| || ||:|| || || | |||::: || || | :||:||
```

```
Db    449 AGACACTCAATGCCAGCCTGTGAAAGCAGCTGGGAGGGAGACTGTACCATGCAAGACCAC 508

Qy   1164 ngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygayytngaygt 1223
            |  || ||::|  | ||:||  | || | : |::| :| ||:  :||:||::| ||:||
Db    509 AGGGGCAGAGCTGCCCAAGACCATGGGAACCCACCTCTTGCATCATCATGACCTGGATGT 568

Qy   1224 nmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacnggntt 1283
            :| ||:|| |  ||: | ||:||:||:|| || :| :| ||:||:| :||  | || ||
Db    569 GAGACATGGAGCCAAAGGAGATCATTTTGGAGCTTTAAGATTTGACTTCCCTGCTGGATT 628

Qy   1284 ymgnacntayatgggnccngt--nccnytntgyttyggncarttyttyccnttyggnacn 1341
            ::  || | :|||||| || ||   || :| | :||:|| ||:||:  :|| ||:|| |
Db    629 TCAGACTTGCATGGGCCCTGTAGCCCCTTTTTTTTTGGCCAATTTCCCCCATTTGGAATG 688

Qy   1342 gcngtntty 1350
            |   |  |:
Db    689 GGGTTAATT 697
```

```
RESULT   9
BI754555/c
LOCUS       BI754555                736 bp    mRNA    linear   EST 25-SEP-2001
DEFINITION  603023384F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5194028 5',
            mRNA sequence.
ACCESSION   BI754555
VERSION     BI754555.1  GI:15746133
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11485  row: g  column: 21
            High quality sequence stop: 733.
FEATURES             Location/Qualifiers
     source          1. .736
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5194028"
                     /clone_lib="NIH_MGC_114"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is  destroyed
                     upon cloning).  Average insert size 1.5 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length  clones and was constructed by C. Gruber
                     (Invitrogen).  Research  Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT      163 a    196 c    185 g    192 t
ORIGIN
```

```
  Query Match             22.3%;  Score 306.6;  DB 10;  Length 736;
  Best Local Similarity   44.1%;  Pred. No. 5.1e-59;
  Matches  306;  Conservative  122;  Mismatches  266;  Indels    0;  Gaps    0;

Qy    659 tngcntggmgnccnmgnmgnaartggttytgyggnacnggnccnggnwsnytntgytgyg 718
          | |  |||:  |   | :   |    | ||: |    || |  |   : :| ||:  :
Db    694 TGGTCTGGACCCAAGGTCCCCATGTTGTGTGCAGCCTAGGACTTGCTGCCTTGTGTCCCA 635
```

```
Qy    719 tncarccnmgngayytngtnccntgygtnccngtnaaywsngcngtngcnwsngarggng 778
          ||  | :|  |::        |          | ::: ||    |  :    :|
Db    634 GCCACTCCAGCCATTGCTAAAAGGGGCCAAGGTACAGCTCAGCCCATGGTTTCAGAGAGT 575

Qy    779 cnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnm 838
          :  ||  |:|| |||||::| || :: || || ||:|| |  || |   |:||::: :
Db    574 GCAAACCCCAACCTTGGCAGCTACCATGTGGTGTTGAGCCTGCGGGTGTACAGAAGTCAA 515

Qy    839 gnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggatgc 898
          | ||:||:|| |||||:|| ||   ::| ||:||:||:|| ||:|  ||: | |||||||
Db    514 GAATTGAGGTTTGGGAACCTCCATCTAGATTTCAGAAGATGTATGAAAATGCCTGGATGC 455

Qy    899 cnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtnc 958
          | :| ||:||:||:|| |  || |  ||  : :: |||:| || :: || :| |  || |
Db    454 CCAGGCAAAAGTTTGCTGCAGGGGAGGGGCCCTCTTGGAGAACCTCTGCTAGGGCAGTGC 395

Qy    959 araarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnm 1018
          |:||:|| ||:|| || |  ||:|| |  ||::|  | || :: || |  |   : :: :
Db    394 AGAAGGGAAATGTGGGGTCAGAGCCCCAACACAGTATCCCTACTGGGGTACTGCCTAGTA 335

Qy   1019 gngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsny 1078
          | || || :| :| :: || || :: :: :  :  ||:|| || :| :  || ||::: :
Db    334 GAGCTGTAAGAAGAAGGCCACCATCCTCCAAACATCAGAATGGTAGATACACTGACAGCT 275

Qy   1079 tncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggna 1138
          |  |  | || || | :|| : :  | ||:|| ||:||:||:||  | ||:|  ||  | |
Db    274 TGTACAAGGTGCCTGGAAAAGCCGCAGACACTCAATGCCAGCCCATGAAAGAAGCCAGCA 215

Qy   1139 tggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtayy 1198
           || : :  | || | :||: |     |  | ||::|  | | :||  | || | : |::
Db    214 GGGGAGCTATGCCCTGCAAAGCCACAGGGGTGGAGCTGCCTACGACTATGGGAACACACC 155

Qy   1199 tnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcny 1258
          | :| ||:  :  :||::| ||: | :| ||:|  || ||: |  |:||:||:|| || :
Db    154 TCTTGCATCACTGTGACTTGGATATGAGACATGCAGTCAAAGGAAATCACTTTGGAGCTT 95

Qy   1259 tnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytntgyttyg 1318
            :  ||:||:||:||  | || ||: | || | :|||||| || ||  | :  |  ||:
Db     94 CAAAATTTGACTGCCCCGCTGGATTTTGGACTTGCATGGGTCCTGTAACCCCTTTGTTTT 35

Qy   1319 gncarttyttyccnttyggnacngcngtnttyac 1352
          | | :||: |:|| ||:|| |  || || ||:||
Db     34 GGCCATTTCTCCCATTTGGAATGGCCGTATTCAC 1

RESULT 10
AG112666
LOCUS       AG112666               654 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-119B01.R, genomic survey sequence.
ACCESSION   AG112666
VERSION     AG112666.1  GI:16733185
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-119B01.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 654)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKSl45
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .654
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-119B01.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      165 a    174 c    180 g    135 t
ORIGIN
```

Query Match 21.8%; Score 300.4; DB 12; Length 654;
Best Local Similarity 49.3%; Pred. No. 1.2e-57;
Matches 296; Conservative 112; Mismatches 191; Indels 2; Gaps 2;

```
Qy    760 gcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 819
          || || || :: ||:|| || :: |  ||:|| |||||::| ||    |  ||  |:||
Db     50 GCTGTAGCTTCAGAGGGTGCAAGCCTCAAGCCTTGGCAGCTTCCACATGATGTTAAGCCT 109

Qy    820 gtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarath 879
          ||    ||  |:||::: :| ||:| :|| |||||:|| ||   ::| ||:||:||:||
Db    110 GTGAATGCACAGAAGTCAAGAATTGGGGTTTGGGAACCTCCACCTAGATTTCAGAAAATG 169

Qy    880 tayggnaaygccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgn 939
          ||:|| ||: | |||||||| :| ||:||:||:||    || || || :: :: |||:|
Db    170 TATGGAAACTCCTGGATGCCCAGGCAGAAGTTTGCTCCTGGGGTGGGGTCCTCATGGAGA 229

Qy    940 acnwsngcnmgngtngtncaraarggnaaygtnggngtgggarccnccncaymgngtnccn 999
          |  :: || :| |  || ||:||:|  ||:|| || ||||||:|| ||  |:::| || ||
Db    230 AACTCTGCTAGGGCAGTGCAGAAGGAAAACGTGGGGTGGGAGCCCCCATACAGAGTCCCT 289

Qy   1000 wsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraar 1059
          :: || || |   : ::  | || || :| :|  : || |  :: :: :| :  ||:||
Db    290 ACTGGGGCACTGCCTAGG-GAGCCGTGAGAAGAGGGCCACTGTCCTCCAGACCCCAGAAT 348

Qy   1060 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgycar 1119
          | :| :: || ||::: :| |  ||:|| || | :||: : || ||:|| ||:  :||
Db    349 CGTAGATCCACTGACAGCTTGCGCCATGTACCTGGAAAAGCCACAGACACTCAACACCAC 408

Qy   1120 ccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacn 1179
          ||  | ||:||  | || |  ||: : || || | :||: |    |  |  |:|  |
Db    409 CCCATGAAGGCAACTGGGAGAGAGGCTGTACCCTGCAAAGCCACAGGGGTGAAACTGCCC 468

Qy   1180 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1239
          ||:||    || | :||::| :| ||:   ||:||::| ||:|| :| ||:||  | ||:
Db    469 AAGACTAGGGGAACCTACCTCTTGCATCAGCATGACCTGGATGTGAGGCACGGAATCAAG 528

Qy   1240 mgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggn 1299
           | ||:||:||:|| || :| :| ||:||:||:|| || || ||::  || | :||||
Db    529 GGAGATCACTTTGGAGCTTTAAGATTTGACTGCCCCACTGGATTTCAAACTTGCATGGGC 588

Qy   1300 ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg 1358
          || ||  || :|  |:| :|| ||:| : |:|| ||:|| |  || |   |:   | :||
Db    589 CCTGTAGCCCCTTAGTTATGGCCAATCTCTCCCATTTGGAATGGCTGCACTTCACCCATG 648

Qy   1359 y 1359
          :
Db    649 T 649
```

```
RESULT 11
AG064424
LOCUS       AG064424               676 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-053H24.R, genomic survey sequence.
```

```
ACCESSION   AG064424
VERSION     AG064424.1  GI:16616226
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-053H24.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 676)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .676
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-053H24.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      164 a    165 c    191 g    156 t
ORIGIN
```

```
  Query Match             21.5%;  Score 296;  DB 12;  Length 676;
  Best Local Similarity   47.9%;  Pred. No. 1.3e-56;
  Matches  294;  Conservative  117;  Mismatches  201;  Indels    2;  Gaps    2;

Qy     760 gcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 819
           ||  | || ::  |:|| || :: |  ||:|| |||||::| ||    |  || ||:|
Db      63 GCCATGGCTTCTAAAGGTGCAAGCCTCAAGCCTTGGCAGCTTCCACATGATGTTGAGCTG 122

Qy     820 gtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarath 879
           |   | ||  |:||::: :| | :| :|| | |||:|| |    ::| ||:||:| :||
Db     123 GCAAGTGCACAGAAGTCAAGAACTGGGGTTTAGGAACCTCTACCTAGATTTCAGAGGATG 182

Qy     880 tayggnaayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgn 939
           ||:|| ||: | ||||||  | :| ||:||:||:|| |  ||    || :: :: |||:|
Db     183 TATGGAAATGCCTGGATGTCCAGGCAGAAGTTTGCCGCAGGGACGGGGTCCTCTTGGAGA 242

Qy     940 acnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccn 999
           || :: || :| |   |   :||:|| ||:|| || | |||: | |  | ::| || ||
Db     243 ACCTCTGCTAGGGCAATGTGGAAGGGAAATGTGGGGTTGGAGTCCCAGCCCAGAGTCCCT 302

Qy    1000 wsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraar 1059
           :: || || || :: ::  | || || :| :| ::  | |  :: :: :| :  ||:||
Db     303 ACTGGGGCACCATCTAGTGGAGCTGTGAGAAG-AGGGCACTGACCTCCAGACCCCAGAAT 361

Qy    1060 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgycar 1119
           || :| :: || ||::: :| || ||:|| || ||:||: :     |:|| ||:  :||:
Db     362 GGTAGATCCACTGACAGCTTACACCATGTGCCTGAAAAAGCTGTAAACACTCAACACCAG 421

Qy    1120 ccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacn 1179
           ||  | ||:||  | || | |||:   || || | :| : |    |  || ||::|  |

Db     422 CCCATGAAAGCATCTGGGAGGGAGGTTGTACCCTGCAGATCCACAGGGGCGGAGCTGCCC 481

Qy    1180 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1239
           ||:||  |.||   : |::| :| ||:     :||::| ||:|| :| ||:|| || ||:
Db     482 AAGACCATGGGAGCCCACCTCTTGCATCAGTGTGACCTGGATGTGAGACATGGAGTCAAA 541

Qy    1240 mgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggn 1299
           | ||:||:||:|  |  :| :| ||:||:||:|   | || ||: | ||   :|||||
Db     542 GGAGATCATTTTGTAGTTTTAAGATTTGACTGCCTCGCTGGATTTTGGACTAGCATGGGG 601

Qy    1300 ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg 1358
           || ||  || :| ||:||:|| ||:||: |:|| ||:|| |  || || ||:|| ||:||
Db     602 CCTGTAGCCCCTTTGTTTTGGCCAATTTCTCCCATTTGGAATGGCTGTATTTACCCAATG 661

Qy    1359 yytntayytncayt 1372
           ::| ||::  ||:|
Db     662 CCTGTACCCCCATT 675
```

```
RESULT  12
AG062447
LOCUS       AG062447                691 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-050M20.F, genomic survey sequence.
ACCESSION   AG062447
VERSION     AG062447.1  GI:16614249
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-050M20.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 691)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-050M20.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      160 a    167 c    201 g    162 t      1 others
ORIGIN
```

```
  Query Match             20.9%;  Score 287.4;  DB 12;  Length 691;
  Best Local Similarity   49.0%;  Pred. No. 1.2e-54;
  Matches  301;  Conservative  110;  Mismatches  196;  Indels    7;  Gaps    3;

Qy     744 ygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggcarytncc 803
           :|| |  |    :   || || || :: ||:|| || :: || ||:|  |||  ::| ||
Db      82 CGTCCGCGGAGCTCAGGCTGTGGCTTCAGAGGGTGCAAGACCCAAGCTTTGGAGGCTTCC 141
```

```
Qy    804 nwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarccnccnat 863
          :: || || ||:|| ||  | ||  |:||::     ||:| :|| |||||:|| ||  |
Db    142 ATGTGGTGTTGAGCCTGTGAGTGCACAGAAGA-----ATCGGGGTTTGGGAACCTCCACT 196

Qy    864 hmgnttycaraarathtayggnaaygccntggatgccnmgncaraarttygcngtnggngt 923
          ::| ||:||:||:|| ||:|  ||: | |||||||| :  ||:||:||:|| || ||
Db    197 TAGATTTCAGAAGATGTATGAAAATACCTGGATGCCCATGCAAAAGTTTGCTGTAGGGTC 256

Qy    924 nggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntgggarcc 983
          ||  : :: |||:| || :: || :| |  ||   :||:|| ||:||  | |  ||: |
Db    257 AGGGCCCTCATGGAGAACTTCTGCTAGGGCAGTGTGGAAGGGAAATGTGAGGTCAGAGGC 316

Qy    984 nccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnws 1043
          || ||::| || || :: || || ||  : ::  | || || :| :|  : || |  ::
Db    317 CCCACACAGAGTCCCTACTGGGGCACCACCCAGTGGAGCTGTGAGAAGAGGGCCACTGTC 376

Qy   1044 nwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnac 1103
          :: :| :| ||:|| || :| :: || ||: : :| || ||:|| || | :||: : ||
Db    377 CTCCAGACTCCAGAATGGTAGATCCACTGATGGCTTGCATCATGTGCCTGGAAAAGCCAC 436

Qy   1104 ngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntayaaracngt 1163
          ||:|| ||:||:||:||  | ||:|| || || | |||: : || || ||:||: |
Db    437 AGACACTCAATGCCAGCCCATGAAAGCAGCCGGGAAGGAGGCTGTACCCTACAAGGCCAC 496

Qy   1164 ngtngcngaryt-nacnaaracngtnggnathtayytnytncaytgycaygayytngayg 1222
          |  || ||::|   | ||:||  | || | : |::  :| | :     :||::| | :|
Db    497 AGGGGCAGAGCTGTTCCAAGACCATGGGAACCCACCCCTTGCCTCAGTGTGACCTGGTTG 556

Qy   1223 tnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacnggnt 1282
          | :|  |:|| || | : | ||:||:| : | || :| :| ||:||:||:    | || |
Db    557 TGAGATATGGAGTCAGAGGAGATCATTGTAGAGCTTTAAGATTTGACTGTGTGGCTGGAT 616

Qy   1283 tymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnttyggnacn 1341
          |: | || | :|||| || ||  |  :| ||:||:|| ||:||: |:|| ||:|  |
Db    617 TTGGGACTTGCATGGGGCCTGTAGCTCCTTTGTTTTGGCCAATTTCTCCCATTTGTAATG 676

Qy   1342 gcngtnttyacnca 1355
          || || ||:|| ||
Db    677 GCTGTATTTACCCA 690


RESULT 13
AG116938
LOCUS       AG116938                720 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-124K01.R, genomic survey sequence.
ACCESSION   AG116938
VERSION     AG116938.1  GI:16737457
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-124K01.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 720)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
            LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-124K01.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      181 a    186 c    187 g    165 t      1 others
ORIGIN


  Query Match             20.8%;  Score 286.2;  DB 12;  Length 720;
  Best Local Similarity   48.5%;  Pred. No. 2.3e-54;
  Matches  298;  Conservative  112;  Mismatches  200;  Indels    4;  Gaps    3;

Qy    760 gcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 819
          ||  | || :: || || || ::  | ||: | |||  ::| ||    ||  | ||:||
Db    110 GCCATGGCTTCAGATGGTGCAAGCACCAAGTCTTGGTGGCTTCCACATGGCATTGAGCCT 169

Qy    820 gtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarath 879
          |  || ||  |:||::: :| ||:||:|| ||| |:|| ||   :   ||:||:|| ||
Db    170 GCGGGTGCACAGAAGTCAAGAATTGAGGTTTGGAAACCTCCACCTGAATTTCAGAACATG 229

Qy    880 tayggnaayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgn 939
          ||:|| ||: | ||||||||| :| ||:||:||:|| |  || |   |    :: |||:|
Db    230 TATGGAAATGCCTGGATGCCCAGGCAGAAGTTTGCTGCAGGGGAGTGGCTCTCATGGAGA 289

Qy    940 acnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccn 999
          || :: || :| |  || ||:||:|| ||:|| || | |||:|| || ||::| || ||
Db    290 ACCTCTGCTAGGGCAGTGCAGAAGGGAAATGTGGGGTCGGAGCCCCCACACAGGGTCCCT 349

Qy   1000 wsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraar 1059
          :: || |   | : ::  | || || :| :|    || |  :: :: :| :  ||:||
Db    350 ACTGGGG-ATCACCTAGTGGAGCTGTGAGAAGACAGCCACTGTCTTCCAGACCCCAGAAT 408

Qy   1060 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgycar 1119
          |  :| :: || ||::: :    :| :|| || | :||: :  | | :|| ||:  :||:
Db    409 GCTAGATCCACTGACACTTGC--ACCTGTGCCTGGAAAAGCTGCAGGTACTCAACATCAG 466

Qy   1120 ccngtnaargcngcnggngatggarwsngtnccntayaaracngtngtngcngarytnacn 1179
          ||  | ||:|| || || | |||: : || || | :||: |    |  |  ||::|  |
Db    467 CCCATGAAAGCAGCTGGGAGGGAGGCTGTACCCTGCAAAGCAACAGAGGTGGAGCTTCCC 526

Qy   1180 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1239
           |:||  | || | :||::| :| ||:     :||::| ||:|| :| ||:||  | ||:
Db    527 GAGACCATGGGAACCTACCTCTTGCATCAGGGTGACCTGGATGTGAGACATGGATTCAAA 586

Qy   1240 mgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggn 1299
          :| ||:||: |:|| || :| :| ||:||: |:|  || || ||: | || | :|||||
Db    587 AGAGATCATATTGGAGCTTTAAGATTTGACAGCCACACTGGATTTTGGACTTGCATGGGC 646

Qy   1300 ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg 1358
          || ||  || :| ||:||:|| ||:||: |:|| ||:|| |  |  |  ||:||  |:||
Db    647 CCTGTAGCCCTTTTGTTTTGGCCAATTTCTCCCATTTGGAATGGTTGCATTTACCTAATG 706

Qy   1359 yytntayytncayt 1372
          ::| ||::  ||:|
Db    707 TCTGTATTCCCATT 720


RESULT 14
AQ037711
LOCUS       AQ037711                623 bp    DNA     linear   GSS 11-JUL-1998
DEFINITION  CIT-HSP-2337G10.TR CIT-HSP Homo sapiens genomic clone 2337G10, DNA
            sequence.
ACCESSION   AQ037711
VERSION     AQ037711.1  GI:3303543
KEYWORDS    GSS.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2337G10.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .623
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2337G10"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      160 a    146 c    172 g    145 t
ORIGIN
```

```
  Query Match             20.6%;  Score 283.8;  DB 12;  Length 623;
  Best Local Similarity   48.1%;  Pred. No. 7.3e-54;
  Matches  286;  Conservative  111;  Mismatches  196;  Indels    2;  Gaps    2;

Qy     784 ccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnath 843
           |  ||:|  |||||::| ||    || || ||: | |  ||  |   :||::: :| | :
Db      26 CTCAAGCTTTGGCAGCTTCCACATGGTGTTGAGTCTGCGGGTACACCAAAGTCAAGAACT 85

Qy     844 gargtntg-ggarccnccnathmgnttycaraarathtayggnaaycentggatgccnmg 902
           ||:|| || |||:|| ||   ::  ||:||:||:|| ||:|| ||: | |||||| | :|
Db      86 GAGGTTTGAGGAACCTCCACCTACATTTCAGAAGATGTATGGAAATGCCTGGATGTCCAG 145

Qy     903 ncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraa 962
            ||:||:||:|| |  || |  ||  : :: |||:| || :: || :| |  || | :||
Db     146 GCAAAAGTTTGCTGCAGGGGCAGGGCCCAGATGGAGAACCTCTGCTAGGGCAGTGCGGAA 205

Qy     963 rggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngc 1022
           :|| ||:|| || | |||:||  | || :| ||  | :: || || |   : ::  | ||
Db     206 GGGAAATGTGGGGTTGGAGCCGTCACAGAGAGTCTCTACTGGGGCACTGCCTAGTGGAGC 265

Qy    1023 ngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytnca 1082
            || :| :|  : ||    :: :: :| :   |:|| || :| ::  | ||::: :| ||
Db     266 TGTGAGAAGAGGGCCATTGTCCTCCAGACCCTAGAATGGTAGATCTGCCGACAGCTTGCA 325

Qy    1083 rcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggnatgga 1142
             |:|   | | :||: : || ||:|| ||:  :||:||  | ||: | ||  | | ||
Db     326 TTATGCAACTGGAAAAGCCACAGACACTCAACACCAGCCCATGAAAACAGCCAGGAGGGG 385

Qy    1143 rwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtayytnyt 1202
           : :  | || | :||: |    |  || ||::   |  |: |  | || | : |:::| :|
Db     386 AGCCATACCCTGCAAAGCCACAGGGGCAGAGTGGCCTGAGGCCATGGGAACCCACCTCTT 445

Qy    1203 ncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmg 1262
            ||:  | :||::  ||:|| :| ||:|| || ||: | ||:||:||:|| || :| :|
Db     446 GCATCAGCGTGACCCAGATGTGAGACATGGTGTCAAAGGAGATCATTTTGGAGCTTTAAG 505
```

```
Qy    1263 nttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgyttyggnc 1321
           ||:||:||:||  | || ||::  || | :| ||| || ||  || :| | :||:|| |
Db     506 ATTTGACTGCCCTGCTGGATTTCAGACTTGTAAGGGGCCAGTAGCCCTTTTCTTTTGGCC 565

Qy    1322 arttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytgyat 1376
           |:||: |:|| ||:|| |  || || ||:|| ||:||::| ||::  ||:||:||
Db     566 AATTTCTCCCATTTGGAATGGCTGTATTTACCCAATGCCTGTATTCCCATTGTAT 620
```

```
RESULT  15
AG097258
LOCUS       AG097258                677 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-098001.R, genomic survey sequence.
ACCESSION   AG097258
VERSION     AG097258.1  GI:16717775
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-098001.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 677)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .677
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-098001.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      178 a    175 c    176 g    147 t      1 others
ORIGIN
```

```
  Query Match             20.6%;  Score 283.2;  DB 12;  Length 677;
  Best Local Similarity   48.8%;  Pred. No. 1e-53;
  Matches  294;  Conservative  113;  Mismatches  187;  Indels    9;  Gaps    3;

Qy     760 gcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 819
           || || || :: ||:||    :: || ||:|| |||||::| || :: || || ||:||
Db      68 GCTGTGGCTTCAGAAGGTAGAAGCCCCAAGCCTTGGCAGCTTCCATGTGGTGTTGAGCCT 127

Qy     820 gtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarath 879
              || ||  |:||::: :  ||:||:|| || ||:|| ||   ::| ||: |:| :||
Db     128 ACAGGTGCACAGAAGTCACAAATTGAGGTTTGTGAACCTCCACCTAGATTT-AGACGATG 186

Qy     880 tayggnaayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgn 939
           ||:|| ||: | |||||||| :|  |:||:||:|| |  ||     |  : :: |||:|
Db     187 TATGGAAACACCTGGATGCCCAGGTAAAAGTTTGCTGCAGGGTGGAGGCCATCATGGAGA 246
```

```
Qy    940 acnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccn 999
          || ::  | :| |  || | :||:|| ||:|  || | |||:|| ||  |::| || ||
Db    247 ACCTCTTCTAGGGCAGTACGGAAAGGAAATGCAGGGTTGGAGCCCCCAAACAGAGTCCCT 306

Qy   1000 wsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraar 1059
          :  || || |   : ::    || || :| :|  : || || :: :: :| :  ||:||
Db    307 AATGGCGCACTGCCTAGTGAAGCTGTAAGAAGAGGACCACCATCCTCCAGACCCCAGAAT 366

Qy   1060 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgycar 1119
          || :| :: ||  |::: :| ||:||:|  || | :||: :  | ||:|| ||: |:||:
Db    367 GGTAGATCCACCTACAGCTTGCAACATGCACCTGGAAAAGCTGCAGACACTCAGCGCCAG 426

Qy   1120 ccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacn 1179
          ||  | ||:|| ||  | | |||::: ||  | | :||: |    |  || ||::|  |
Db    427 CCCATGAAAGCAGCCAGGAGGGAGACTGTCTCCTGCAAATCCACAGGGGCAGAGCTGCCC 486

Qy   1180 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1239
          ||:||  | || | : |::| :| ||:   | :||::| ||:|| :| ||:|| || ||:
Db    487 AAGACCATGGGAACCCACCTCTTGCATCAACGTGACCTGGATGTGAGGCATGGAGTCAAA 546

Qy   1240 mgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggn 1299
           | ||:||:| :|| ||       ||:||:||:||  | || ||: | || | :|||||
Db    547 GGAGATCATTNTGGAGC-------TTTGATTGCCCTGCTGGATTTTGGACTTGCATGGGC 599

Qy   1300 ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg 1358
          || ||  || :| | :||:|  ||:||: |:|| ||:|  |  || || ||:|| ||:||
Db    600 CCTGTATCCCCTGTATTTTGAGCAATTTATCCCATTTGAAATGGCTGTATTTACCCAATG 659

Qy   1359 yyt 1361
          ::|
Db    660 CCT 662


Search completed: July  8, 2002, 18:50:56
Job time: 19336 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 18:08:58 ; Search time 5686.68 Seconds
(without alignments)
6369.944 Million cell updates/sec

Title: US-09-997-610-7
Perfect score: 1731
Sequence: 1 atgytnathgargcntaygg.....gyytntayytncaytgyatg 1731

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	921	53.2	145880	9	HS302D9	Z82198 Human DNA s
2	497.8	28.8	154090	9	AC025577	AC025577 Homo sapi
3	497.8	28.8	186660	2	AC026107	AC026107 Homo sapi
c 4	493.6	28.5	124518	9	AL138962	AL138962 Human DNA
5	493	28.5	123070	2	AC008799	AC008799 Homo sapi
c 6	493	28.5	189768	2	AC044889	AC044889 Homo sapi
7	490.4	28.3	198545	2	AC017063	AC017063 Homo sapi
c 8	488.8	28.2	161575	2	AC025233	AC025233 Homo sapi
9	488.8	28.2	163674	2	AC087500	AC087500 Homo sapi
c 10	488.8	28.2	175639	2	AC015727	AC015727 Homo sapi
c 11	488.6	28.2	181842	2	AL391823	AL391823 Homo sapi
12	488.2	28.2	77405	2	AL353634	AL353634 Homo sapi
c 13	488.2	28.2	124531	9	HSJ633H17	AL049710 Human DNA
14	487.6	28.2	40714	9	HSU212C1	Z69722 Human DNA s
c 15	487.6	28.2	159712	9	AP004219	AP004219 Homo sapi
c 16	487	28.1	125295	2	AL672061	AL672061 Homo sapi
c 17	487	28.1	146743	2	AC093588	AC093588 Homo sapi
c 18	487	28.1	152544	9	CNS05TEJ	AL359232 Human chr
c 19	487	28.1	166679	9	AC079899	AC079899 Homo sapi
c 20	487	28.1	174662	2	AC026036	AC026036 Homo sapi
c 21	487	28.1	176426	9	AC007370	AC007370 Homo sapi
c 22	487	28.1	203726	2	AC011882	AC011882 Homo sapi
23	486	28.1	139005	2	AF286112	AF286112 Homo sapi
c 24	486	28.1	156458	2	AC025647	AC025647 Homo sapi
c 25	486	28.1	165901	9	AP003474	AP003474 Homo sapi
c 26	486	28.1	171427	9	AC021590	AC021590 Homo sapi
c 27	485.8	28.1	99084	2	AC026420	AC026420 Homo sapi
c 28	485.8	28.1	133790	9	AC010280	AC010280 Homo sapi
c 29	485.8	28.1	150332	9	AC004921	AC004921 Homo sapi
30	485.8	28.1	174293	9	AC008816	AC008816 Homo sapi
31	485.8	28.1	203043	9	AC017093	AC017093 Homo sapi
32	485.4	28.0	131215	9	AC079614	AC079614 Homo sapi
33	485	28.0	107885	9	AC006389	AC006389 Homo sapi
34	484.6	28.0	173480	9	CNS00M8T	AL079343 Human chr
c 35	483	27.9	164370	9	AC097460	AC097460 Homo sapi
36	482.6	27.9	112515	9	AL139216	AL139216 Human DNA
37	482.6	27.9	142224	2	AC096535	AC096535 Homo sapi
38	482.6	27.9	178535	2	AL354978	AL354978 Homo sapi
c 39	482	27.8	168502	9	AC091005	AC091005 Homo sapi
40	481.6	27.8	38235	9	AC004559	AC004559 Homo sapi
41	481.4	27.8	234622	2	AC110089	AC110089 Homo sapi
c 42	479.8	27.7	187883	9	AC008268	AC008268 Homo sapi
43	479.6	27.7	167891	9	AC013439	AC013439 Homo sapi
c 44	479.2	27.7	73390	9	AC012038	AC012038 Homo sapi
45	479.2	27.7	94730	9	AP000230	AP000230 Homo sapi

ALIGNMENTS

RESULT 1
HS302D9
LOCUS HS302D9 145880 bp DNA linear PRI 12-DEC-1999
DEFINITION Human DNA sequence from clone RP1-302D9 on chromosome 22 Contains GSSs, complete sequence.
ACCESSION Z82198
VERSION Z82198.2 GI:6572207
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 145880)
AUTHORS Bridgeman,A.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Dec 13, 1999 this sequence version replaced gi:3164067.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22
RP1-302D9 is from the library RPCI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP1-302D9 The true left end of clone CTA-282F2 is at 69682 in this sequence. The true right end of clone CTA-415G2 is at 55167 in this sequence.

FEATURES Location/Qualifiers
source 1. .145880
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone="RP1-302D9"
/clone_lib="RPCI-1"
repeat_region 188. .245
/note="MER3 repeat: matches 144. .209 of consensus"
repeat_region 246. .571
/note="AluSx repeat: matches 1. .312 of consensus"
repeat_region 572. .759
/note="MER3 repeat: matches 1. .144 of consensus"
repeat_region 783. .933
/note="MER5A repeat: matches 26. .187 of consensus"
repeat_region 1033. .1336
/note="AluSp repeat: matches 1. .299 of consensus"
repeat_region 1450. .1583
/note="MIR repeat: matches 24. .160 of consensus"
repeat_region 1687. .1752
/note="L2 repeat: matches 2593. .2661 of consensus"
repeat_region 2350. .2660
/note="AluSc repeat: matches 3. .309 of consensus"
repeat_region 2684. .2981
/note="AluSq repeat: matches 2. .300 of consensus"
repeat_region 3323. .3343
/note="MLT1E repeat: matches 116. .136 of consensus"
repeat_region 3344. .3652
/note="AluY repeat: matches 1. .309 of consensus"
repeat_region 3653. .3928
/note="MLT1E repeat: matches 136. .359 of consensus"
repeat_region 3929. .4278
/note="THE1B repeat: matches 3. .364 of consensus"
repeat_region 4279. .4485
/note="MLT1E repeat: matches 359. .568 of consensus"
repeat_region 5073. .5176
/note="52 copies 2 mer ct 78 conserved"
repeat_region 5181. .5491
/note="AluJb repeat: matches 1. .311 of consensus"
repeat_region 6369. .6485
/note="L2 repeat: matches 2579. .2705 of consensus"
repeat_region 6647. .6685
/note="MADE1 repeat: matches 1. .23 of consensus"
repeat_region 6686. .6987
/note="AluSx repeat: matches 1. .302 of consensus"
repeat_region 6988. .7036
/note="MADE1 repeat: matches 23. .77 of consensus"
repeat_region 7482. .7754
/note="AluJb repeat: matches 9. .290 of consensus"
repeat_region 7775. .8060
/note="AluJo repeat: matches 1. .295 of consensus"
repeat_region 8414. .8551
/note="L2 repeat: matches 2553. .2706 of consensus"
repeat_region 8914. .9030
/note="MIR repeat: matches 147. .262 of consensus"
repeat_region 9110. .9280
/note="MIR repeat: matches 91. .262 of consensus"
repeat_region 9283. .9412
/note="MIR repeat: matches 15. .144 of consensus"
repeat_region 9521. .9679
/note="FAM repeat: matches 3. .161 of consensus"
repeat_region 9820. .10225
/note="MSTB repeat: matches 2. .425 of consensus"
misc_feature complement(10179. .10678)
/note="match: GSS: Em:B56592"
misc_feature complement(10204. .10728)
/note="match: GSS: Em:AQ701486"
misc_feature complement(10249. .10706)
/note="match: GSS: Em:AQ225495"
repeat_region 10312. .10383
/note="MIR repeat: matches 79. .150 of consensus"
misc_feature 10718. .11310
/note="match: GSS: Em:B14024"
misc_feature 10784. .11201
/note="match: GSS: Em:B43656"
repeat_region 11838. .11946
/note="MIR repeat: matches 20. .137 of consensus"
repeat_region 12174. .12445
/note="L2 repeat: matches 1988. .2275 of consensus"
repeat_region 12444. .12642
/note="MIR repeat: matches 63. .241 of consensus"
misc_feature 13017. .13369
/note="match: STS: Em:G49301"
repeat_region 13331. .13397
/note="MIR repeat: matches 174. .244 of consensus"
repeat_region 13398. .13698
/note="AluSp repeat: matches 1. .302 of consensus"
repeat_region 13699. .13810
/note="MIR repeat: matches 76. .174 of consensus"
repeat_region 13806. .13919
/note="MIR repeat: matches 77. .189 of consensus"
repeat_region 13945. .14060
/note="MIR repeat: matches 24. .142 of consensus"
repeat_region 14061. .14367
/note="AluY repeat: matches 1. .301 of consensus"
repeat_region 14368. .14452
/note="MIR repeat: matches 141. .225 of consensus"
repeat_region 14589. .14679
/note="MIR repeat: matches 173. .262 of consensus"
misc_feature 14597. .15201
/note="match: GSS: Em:AQ553482"
misc_feature 14616. .15060
/note="match: GSS: Em:AQ370601"
repeat_region 14868. .15040
/note="MIR repeat: matches 49. .233 of consensus"
repeat_region 15071. .15188
/note="L2 repeat: matches 2112. .2239 of consensus"
repeat_region 15304. .15399
/note="MLT1B repeat: matches 1. .99 of consensus"
repeat_region 15490. .15662
/note="AluSg1 repeat: matches 2. .114 of consensus"
repeat_region 15669. .15727
/note="MLT1B repeat: matches 119. .178 of consensus"
repeat_region 15728. .16027
/note="AluSc repeat: matches 1. .299 of consensus"
repeat_region 16028. .16245
/note="MLT1B repeat: matches 178. .390 of consensus"
repeat_region 16546. .16854
/note="AluY repeat: matches 1. .300 of consensus"
repeat_region 18296. .18323
/note="MSTA repeat: matches 2. .29 of consensus"
repeat_region 18324. .18392

```
                     /note="MER66-internal repeat: matches 4919. .4993 of
                     consensus"
     repeat_region   18393. .18712
                     /note="AluJb repeat: matches 1. .311 of consensus"
     repeat_region   18713. .19133
                     /note="MER66-internal repeat: matches 4548. .4919 of
                     consensus"
     misc_feature    complement(18872. .19230)
                     /note="match: GSS: Em:AQ005063"
     misc_feature    19251. .19719
                     /note="match: GSS: Em:B1417g"
     repeat_region   19537. .20290
                     /note="HERVFH21 repeat: matches 4657. .5784 of consensus"
     repeat_region   20317. .20382
                     /note="33 copies 2 mer ta 68 conserved"
     repeat_region   20513. .20666
                     /note="77 copies 2 mer tt 70 conserved"
     repeat_region   20682. .21008
                     /note="AluSg1 repeat: matches 1. .306 of consensus"
     repeat_region   21239. .21553
                     /note="HUERS-P3 repeat: matches 4410. .4713 of consensus"
     repeat_region   21882. .22254
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   22302. .22537
                     /note="MER66-internal repeat: matches 2186. .2417 of
                     consensus"
     repeat_region   22538. .22850
                     /note="AluSp repeat: matches 1. .313 of consensus"
     repeat_region   22851. .23801
                     /note="MER66-internal repeat: matches 1210. .2186 of
                     consensus"
     repeat_region   23905. .23989
                     /note="MER66-internal repeat: matches 3017. .3102 of

  Query Match             53.2%;  Score 921;  DB 9;  Length 145880;
  Best Local Similarity   58.2%;  Pred. No. 7.9e-215;
  Matches  750;  Conservative  300;  Mismatches  239;  Indels    0;  Gaps    0;

Qy     443 tnggnccnccnggngcnccngngnytnccncartayacnggngarathwsngaratgacna 502
           | |     || || || || || :| || ||:||:|| || ||:||::: ||:|||||  |
Db   37254 TGGAATGTCCTGGTGCCCCAGGTTTACCACAATATACAGGAGAAATAAGTGAAATGACAA 37313

Qy     503 artgyccntgyccngayathgarmgnwsngcnttyacngtnaarytnwsnggnaarytnc 562
           |:||:|| ||:|| ||:||:||::| :: || ||:|| || ||::| :: || ||::| |
Db   37314 AATGCCCCTGTCCTGATATAGAAAGGTCAGCCTTTACTGTGAAGCTCAGTGGAAAACTTC 37373

Qy     563 cnytnccnttyaarccnathathttyacnggngtnytntayaaygcncarmgngayytna 622
           | :| || ||:||:|| ||:||:||:|| || || :| ||:||:|| ||::| ||::| |
Db   37374 CTCTTCCTTTCAAGCCCATCATCTTCACAGGGGTCCTGTACAATGCCCAGAGGGATTTAA 37433

Qy     623 argargcnatgggngtnttygcntgymgngtnccnggnaaytaytaywsnwsnttygayg 682
           |:||:|| ||||| || ||:|| ||::| || || || ||:||:||::: :: ||:||:|
Db   37434 AGGAGGCCATGGGAGTCTTTGCTTGCAGGGTGCCTGGGAATTACTACTCCAGCTTTGATG 37493

Qy     683 tngarytncaycaytgyaargtnaayathtggytnatgmgnaarcarathytngcnaaya 742
           | ||::| ||:||:||:||:|| ||:||:|||:| |||:| ||:||:||::| || ||:|
Db   37494 TTGAGCTGCATCATTGCAAGGTGAATATTTGGCTAATGAGGAAGCAAATTTTGGCTAATA 37553

Qy     743 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnt 802
           |:||:||:||::: ||:||:||::: ||:||:||:|| || ||||| :| :| ||:|| |
Db   37554 AGGAAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGGCAT 37613

Qy     803 tywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtna 862
           |::: ||:||::| ||:|| ||:||:||::: :: ||:||::| ||:|| ||:||:|| |
Db   37614 TCAGTTTCATAAGGGAGGCAGAGCATAAGAGTTCAGAAAATTTGCACCCTGACAATGTGA 37673

Qy     863 thaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyatht 922
           |:||:||:||:||:|| ||::: ||:|| ||:||:||::| || || ||:||:||:||:|
Db   37674 TAAAAAAGAAAAACCCATTTTCTGAGGGGAAATTCAAGCTGGCTGCAGAAATTTGCATAT 37733

Qy     923 gyaaygargarytnaaygtnaaycncargayaayggngaraayathwsntggacntgyc 982
           |:||:||:||::| ||:|| ||:|| ||:||:||:|| ||:||:||::: ||||| ||:|
Db   37734 GTAATGAGGAGCTGAATGTTAATCCTCAAGACAATGGGGAAAATATCTCCTGGACATGTC 37793

Qy     983 armgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttyt 1042
           |::| :: :: ||:||::: ||:||::: :| || |||:| || :| :| ||:|||||:|
Db   37794 AGAGGTCTTCACAGCAGTCCATCAAATCACTGGCCTGGAGGCCTAGGAGAAAATGGTTTT 37853

Qy    1043 gyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtnc 1102
           |:|| || || || || :: :| ||:||:|| ||:|| :| ||::| || || ||:|| |
Db   37854 GTGGGACAGGCCCAGGGTCCCTGTGCTGTGTGCAGCCTAGAGACTTGGTGCCCTGTGTCC 37913

Qy    1103 cngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsng 1162
           | || ||::: || || || :: ||:|| || :: || ||:|| ||||||:| || :: |
Db   37914 CAGTTAATTCAGCTGTGGCTTCAGAGGGTGCAAGCCCCAAGCCTTGGCAGCTTCCAAGTG 37973

Qy    1163 gngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnt 1222
           | || ||:|| || || || ||:||::: :| ||:||:|| |||||:|| || ||::| |
Db   37974 GTGTTGAGCCTGTGGGTGCAAAGAAGTCAAGAATTGAGGTTTGGGAACCTCCAATCAGAT 38033

Qy    1223 tycaraarathtayggnaayccntggatgccnmgncaraarttygcngtnggngtnggnw 1282
           |:||:||:||:||:|| ||:|| |||||||| :| ||:||:||:|| || || || || :
Db   38034 TTCAGAAGATATATGGAAACCCCTGGATGCCCAGGCAGAAGTTTGCTGTAGGGGTGGGGT 38093

Qy    1283 snwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccnc 1342
           : :: |||:| || :: || :| || || ||:||:|| ||:|| || |||||:|| || |
Db   38094 CCTCATGGAGAACCTCTGCAAGGGTAGTACAAAAGGGAAATGTTGGGTGGGAGCCCCCAC 38153

Qy    1343 aymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnm 1402
           |:| || || :: || || || :: :: :| || || :| :| :: || || :: :: :
Db   38154 ACAGAGTCCCCAGTGGGGCTCCATCTAGTAGAGCTGTGAGAAGAAGTCCACCATCCTCCA 38213

Qy    1403 gnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngaya 1462
           | :| ||:||:|| :| :: || ||::: :| ||:||:|| || ||:||::: || ||:|
Db   38214 GACTCCAGAAGGGTAGATCCACTGACAGCTTGCAGCATGTGCCTGAAAAATCCACAGACA 38273

Qy    1463 cncartgycarccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtng 1522
           | ||:||:||:|| || ||:|| || || |||||::: || || ||:||:|| || || |
Db   38274 CTCAGTGCCAGCCTGTGAAAGCAGCAGGGATGGAGTCTGTACCCTACAAAACCGTAGTGG 38333

Qy    1523 cngarytnacnaaracngtnggnathtayytnytncaytgycaygayytngaygtnmgnc 1582
           | ||::| || ||:|| || || ||:||::| :| ||:||:||:||::| ||:|| :| |
Db   38334 CAGAGCTGACCAAGACCGTGGGAATCTACCTCTTGCATTGTCATGACCTGGACGTGAGAC 38393

Qy    1583 ayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgna 1642
           |:|| || ||::| ||:||:||:|| || :| :| ||:||:||:|| || || ||::| |
Db   38394 ATGGAGTCAAAAGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCCACTGGATTTCGGA 38453

Qy    1643 cntayatgggnccngtnccnytntgyttyggncarttyttyccnttyggnacngcngtnt 1702
           | ||:|||| || || || :| ||:||:|| ||:||:||:|| ||:|| || || || |
Db   38454 CTTATATGGGGCCCGTACCCCTTTGTTTTGGCCAATTTTTTCCATTTGGAACTGCCGTAT 38513

Qy    1703 tyacncartgyytntayytncaytgyatg 1731
           |:|| ||:|||::| ||::| ||:||:|||
Db   38514 TTACCCAATGCCTGTACCTCCATTGTATG 38542


RESULT   2
AC025577
LOCUS       AC025577              154090 bp    DNA     linear   PRI 25-AUG-2000
DEFINITION  Homo sapiens 12 BAC RP11-13C3 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC025577
VERSION     AC025577.15  GI:9910028
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154090)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
```

Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Hohnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Muzny,D., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Nelson,D. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 154090)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 154090)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Aug 25, 2000 this sequence version replaced gi:9664948.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

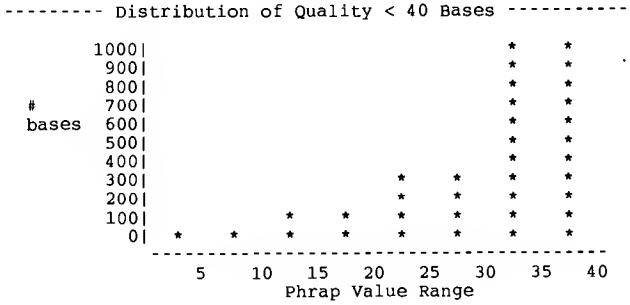
---------- Summary Statistics ----------

Contig length:	154090
Phrap values in estimate:	153545
Average error rate (BCM-Phrap estimate):	0.000261755
Fraction of Phrap values less than 40 :	0.0284086
Number of consensus changing edits:	22
Number of N's in consensus :	0

---------- Consensus changing edits ----------

Position	Original+Context	Edited+Context
9347	aacacagacc(n)ttttttttt	aacacagacc(t)ttttttttt
10135	atttgcccct(n)taggaaaaga	atttgcccct(t)taggaaaaga
10353	ctgtcaccag(n)aaaagatcca	ctgtcaccag(t)aaaagatcca
10417	agcaatgcag(n)ggctacagaa	agcaatgcag(t)ggctacagaa
61959	ctttactatt(n)tgtttattct	ctttactatt(t)gttttattct
61960	tttactattn(t)gtttattcta	tttactattt(g)ttttattcta
61961	ttactattnt(g)tttattctaa	ttactatttg(t)tttattctaa
61982	acaacacaag(n)ntaggtttgg	acaacacaag(a)ttaggtttgg
61983	caacacaagn(n)taggtttggn	caacacaaga(t)taggtttggt
61993	ntaggtttgg(n)ttactcttcc	ttaggtttgg(t)ttactcttcc
62205	tcattgaccc(n)ctggtcattc	tcattgaccc(a)ctggtcattc
62956	actgcaacct(n)tgcctcccag	actgcaacct(c)tgcctcccag
63722	tttacatata(n)cagtacagta	tttacatata(a)cagtacagta
90212	cctagaaaaa(n)gacnttttct	cctagaaaaa(t)gacttttttct
90216	gaaaaangac(n)ttttcttttt	gaaaaatgac(t)ttttcttttt
90228	tttcttttta(n)ttttttcnat	tttcttttta(t)ttttttctat
90236	tantttttc(n)atctttaggg	tatttttttc(t)atctttaggg
90443	accttagatg(n)ctcctcccag	accttagatg(a)ctcctcccag
91405	cacagcttaa(n)caaggtaaga	cacagcttaa(a)caaggtaaga
122290	aaaaaaagaa(n)gtaagagaaa	aaaaaaagaa(a)gtaagagaaa
137090	acagagaaag(n)caaaacactc	acagagaaag(a)caaaacactc
141632	tgaagccact(n)gaaaagtaat	tgaagccact(t)gaaaagtaat

--------- Distribution of Quality < 40 Bases ----------



Version: 1.01 qxfo.

FEATURES Location/Qualifiers

```
     source          1. .154090
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-13C3"
     repeat_region   54. .288
                     /rpt_family="MIR"
     repeat_region   1020. .1146
                     /rpt_family="MER3"
     repeat_region   complement(1248. .1584)
                     /rpt_family="MER46B"
     repeat_region   complement(1624. .1682)
                     /rpt_family="MIR"
     repeat_region   2630. .2761
                     /rpt_family="GA-rich"
     repeat_region   4088. .4216
                     /rpt_family="L2"
     repeat_region   5232. .5301
                     /rpt_family="MER102"
     repeat_region   complement(5450. .5749)
                     /rpt_family="AluY"
     repeat_region   6571. .6616
                     /rpt_family="(TG)n"
     repeat_region   complement(6678. .7320)
                     /rpt_family="L1ME3"
     repeat_region   7419. .7454
                     /rpt_family="AT_rich"
     repeat_region   complement(7741. .7807)
                     /rpt_family="MER5A"
     repeat_region   8113. .8267
                     /rpt_family="MER5A"
     repeat_region   8385. .8587
                     /rpt_family="MIR"
     repeat_region   8589. .8630
                     /rpt_family="L2"
     repeat_region   8666. .8695
                     /rpt_family="(TTA)n"

  Query Match             28.8%;  Score 497.8;  DB 9;  Length 154090;
  Best Local Similarity   47.7%;  Pred. No. 1.6e-110;
  Matches  488;  Conservative  216;  Mismatches  282;  Indels   36;  Gaps    2;

Qy     744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
           :||:||:||::: ||: |:||::: ||: |:||:|| || | | | :  :| ||:|  ||
Db 101633 GGAAGAAATTTCTAAGTAGCAAAGCATTAAAGAGGTGACTTAGATGTGGTTAAAGGTATT 101692

Qy     804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
           ::: | :||::| ||:|| ||:||:||: : :  ||:||::| || || ||:||:|| ||
Db 101693 CAGTTATATAAGGGAGGCAGAGCATAAACGTTTGGAAAATTTGCAGCCTGACAATGTGAT 101752

Qy     864 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
           : |:||:||:||:|| ||::: ||:|| || || ||::| || || ||:||:||:||: |
Db 101753 AGAAAAGAAAAACCCATTTTCTGAGGGGAATTTGAAGCTGGCTGCAGAAATTTGCATAAG 101812

Qy     924 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 983
           :||: |:||::  ||:|| ||:|| ||:||:||:|| ||:||: |:::   | | | :||
Db 101813 TAACAAGGAGCCTAATGTTAATCCCCAAGACAATGGGGAAAATGTCTCCAAGGCATATCA 101872

Qy     984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 1043
           ::| :: ::   :||: : ||:  ::: :  ||  ||:| || :| :| ||:|||||:||
Db 101873 GAGGTCTTCATGGCAGCCCATCCCATCACAGGCCCGGAGGCCTAGGAGAAAATGGTTTTG 101932

Qy    1044 yggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt--- 1100
           : |  | || || ||  : :  ||:||:|| ||:|| :| ||::| || || ||:||
Db 101933 TAGGCCAGGCCCAGGAGCCCCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGTGTATA 101992

Qy    1101 ------------------------------nccngtnaaywsngcngtngcnwsngar 1128
                                           |       ::: || || || :: ||:
Db 101993 CCAGCTGCTGCAGTGGTGGCTGAAAGGGGCCAGCATAGAGCTTGGGCCGTGGCTTCGGAG 102052

Qy    1129 ggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtngngngcnaaraar 1188
           || || :: || ||:|| |||||::| || :: |  || ||:|| |  || ||  |:||:
Db 102053 GGTGCAAGCCCCAAGCCTTGGCAATTTCCATGTGTTGTTGAGCCTGAGGGTGCACAGAAG 102112

Qy    1189 wsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaaycentgg 1248
           :: :| ||:||:|  |||||:|| ||   ::| ||:||:| :|| ||:|| ||: | |||
Db 102113 TCAAGAATTGAGGCTTGGGAGCCTCCACCTAGATTTCAGAGGATGTATGGAAATGCCTGG 102172

Qy    1249 atgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtn 1308
           ||| | :| ||:||:||:||    || || ||  : :: |||:| || :: || :| |
Db 102173 ATGTCCAGGCAGAAGTTTGCTTCAGGGGTGGGGCCCTCATGGAGAACCTCTGCTAGGGCA 102232

Qy    1309 gtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsn 1368
           || ||:||:|| ||:|| || | |||:||    ||::| || || :: || || || ::
Db 102233 GTGCAGAAGGGAAATGTGGGGTTGGAGCCCATACACAGAGTCCCTACTGGGGCACCATCT 102292

Qy    1369 wsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngay 1428
           ::  | || || :| :|  : || || :: :: :| :  ||:|| || :| :: ||  |:
Db 102293 AGGGGAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAATGGTAGATCCACCAAC 102352

Qy    1429 wsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcn 1488
           :: :| || ||:|| || | :||: : || || || ||:  :||:||  | ||:|| ||
Db 102353 AGCTTGCACCATGTACCTGGAAAAGCCACAGAGACTCAACACCAGCCCATGAAAGCAGCC 102412

Qy    1489 ggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1548
              | |||: : || || | :||: |     |  || ||::|  | ||:||  | || | :
Db 102413 AAGAGGGAGCCTGTACCCTGCAAAGCCACAGAAGCAGAGCTTCCCAAGACCATGGGAACC 102472

Qy    1549 tayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggn 1608
            |::| :| ||:     :||::| ||:|| :  ||:|  || ||: | ||:||: |: |
Db 102473 CACCTCTTGCATCAGTGTGACCTAGATGTGAAACATGCAGTCAAAGGAGATCATATTAGA 102532

Qy    1609 gcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntg 1667
            | :| :| ||:||:| :||  | || ||::  || | :|||||  |   |  || :| ||
Db 102533 ACTTTGAGATTTGACTACCCTTCTGGGTTTCAGACTTGCATGGGGCATATAGCCCCTTTG 102592

Qy    1668 yttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1727
           :||:|| ||:||: |:|| ||: | || || || ||:|| ||:||::| ||::  ||:||
Db 102593 TTTTGGCCAATTTCTCCCATTTAGAACAGCTGTATTTACTCAATGCCTGTATCGCCATTG 102652

Qy    1728 ya 1729
           :|
Db 102653 CA 102654


RESULT   3
AC026107
LOCUS       AC026107              186660 bp    DNA     linear   HTG 06-FEB-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-307L1, WORKING DRAFT
            SEQUENCE, 1 unordered pieces.
ACCESSION   AC026107
VERSION     AC026107.22  GI:13899178
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186660)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
```

```
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186660)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 1, 2001 this sequence version replaced gi:13877175.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HAMK
                Center clone name: RP11-307L1
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 18% of reads
                Chemistry: Dye-terminator Big Dye: 82% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 186630 bases at least Q40
                Consensus quality: 186659 bases at least Q30
                Consensus quality: 186660 bases at least Q20
                Estimated insert size: 187779; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 10.7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   186660: contig of 186660 bp in length.
FEATURES            Location/Qualifiers
     source         1. .186660
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="12"
                    /clone="RP11-307L1"
BASE COUNT    57089 a  37346 c  35962 g  56263 t
ORIGIN
```

```
  Query Match           28.8%;  Score 497.8;  DB 2;  Length 186660;
  Best Local Similarity  47.7%;  Pred. No. 1.8e-110;
  Matches  488;  Conservative  216;  Mismatches  282;  Indels   36;  Gaps    2;

Qy    744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
          :||:|||:||::: ||: |:||::: ||: |:||:|| || | | | :  :| ||:|  ||
Db  71882 GGAAGAAATTTCTAAGTAGCAAAGCATTAAAGAGGTGACTTAGATGTGGTTAAAGGTATT 71941

Qy    804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
          ::: | :||:|  |:|| ||:||:||: : :  ||:||::| || || ||:||:|| ||
Db  71942 CAGTTATATAAGGAAGGCAGAGCATAAACGTTTGGAAAATTTGCAGCCTGACAATGTGAT 72001

Qy    864 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
          : |:||:||:||:|| ||::: ||:|| || || ||::| || || ||:||:||:||: |
Db  72002 AGAAAAGAAAAACCCATTTTCTGAGGGGAATTTGAAGCTGGCTGCAGAAATTTGCATAAG 72061

Qy    924 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 983
          :||: |:||::  ||:|| ||:|| ||:||:||:|| ||:||: |:::   | | | :||
Db  72062 TAACAAGGAGCCTAATGTTAATCCCCAAGACAATGGGGAAAATGTCTCCAAGGCATATCA 72121

Qy    984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 1043
          ::| :: ::   :||: : ||:  ::: :  ||  ||:| || :| :| ||:|||||:||
Db  72122 GAGGTCTTCATGGCAGCCCATCCCATCACAGGCCCGGAGGCCTAGGAGAAAATGGTTTTG 72181

Qy   1044 yggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt--- 1100
          : |  | || || ||  : :  ||:||:|| ||:|| :| ||::| || || ||:||
Db  72182 TAGGCCAGGCCCAGGAGCCCCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGTGTATA 72241

Qy   1101 ------------------------------nccngtnaaywsngcngtngcnwsngar 1128
                                         |       ::: || || || :: ||:
Db  72242 CCAGCTGCTGCAGTGGTGGCTGAAAGGGGCCAGCATAGAGCTTGGGCCGTGGCTTCGGAG 72301

Qy   1129 ggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraar 1188
          || || :: || ||:|| |||||::| || :: || || ||:|| |  || ||  |:||:
Db  72302 GGTGCAAGCCCCAAGCCTTGGCAATTTCCATGTGGTGTTGAGCCTGAGGGTGCACAGAAG 72361

Qy   1189 wsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaaycentgg 1248
          :: :| ||:||:|  |||||:|| ||   ::| ||:||:| :|| ||:|| ||: | |||
Db  72362 TCAAGAATTGAGGCTTGGGAGCCTCCACCTAGATTTCAGAGGATGTATGGAAATGCCTGG 72421

Qy   1249 atgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtn 1308
          ||| | :| ||:||:||:||    || || ||  : :: |||:| || :: || :| |
Db  72422 ATGTCCAGGCAGAAGTTTGCTTCAGGGGTGGGGCCCTCATGGAGAACCTCTGCTAGGGCA 72481

Qy   1309 gtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsn 1368
          || ||:||:|| ||:|| || | |||:||    ||::| || || :: || || || ::
Db  72482 GTGCAGAAGGGAAATGTGGGGTTGGAGCCCATACACAGAGTCCCTACTGGGGCACCATCT 72541

Qy   1369 wsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngay 1428
          ::  | || || :| :|  : || || :: :: :| :  ||:|| || :| :: ||  |:
Db  72542 AGGGGAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAATGGTAGATCCACCAAC 72601

Qy   1429 wsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcn 1488
          :: :| || ||:|| || | :||: : || || || ||:  :||:||  | ||:|| ||
Db  72602 AGCTTGCACCATGTACCTGGAAAAGCCACAGAGACTCAACACCAGCCCATGAAAGCAGCC 72661

Qy   1489 ggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1548
             | |||: : || || | :||: |    |  || ||:::|  | ||:||  | || | :
Db  72662 AAGAGGGAGCCTGTACCCTGCAAAGCCACAGAAGCAGAGCTTCCCAAGACCATGGGAACC 72721

Qy   1549 tayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggn 1608
           |::| :| ||:     :||::| ||:|| :  ||:|  || ||: | ||:||: |: |
Db  72722 CACCTCTTGCATCAGTGTGACCTAGATGTGAAACATGCAGTCAAAGGAGATCATATTAGA 72781

Qy   1609 gcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntg 1667
           | :| :| ||:||:| :||  | || ||::  || | :|||| |   |  || :| ||
Db  72782 ACTTTGAGATTTGACTACCCTTCTGGGTTTCAGACTTGCATGGGGCATATAGCCCCTTTG 72841

Qy   1668 yttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1727
          :||:|| ||:||: |:|| ||: | :|| || || ||:|| ||:||::| ||::  ||:||
Db  72842 TTTTGGCCAATTTCTCCCATTTAGAACAGCTGTATTTACTCAATGCCTGTATCGCCATTG 72901

Qy   1728 ya 1729
```

```
          :|
Db  72902 CA 72903


RESULT   4
AL138962/c
LOCUS       AL138962              124518 bp    DNA     linear   PRI 28-JAN-2001
DEFINITION  Human DNA sequence from clone RP11-279N8 on chromosome 13, complete
            sequence.
ACCESSION   AL138962
VERSION     AL138962.16  GI:12597036
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124518)
  AUTHORS   Bates,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 29, 2001 this sequence version replaced gi:12584354.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-279N8 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-279N8 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true right end of clone RP11-279N8 is at 124518 in this
            sequence. The true left end of clone RP11-528D24 is at 114969 in
            this sequence. The true right end of clone RP11-214O11 is at 100 in
            this sequence.
FEATURES             Location/Qualifiers
     source          1. .124518
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-279N8"
                     /clone_lib="RPCI-11.1"
     misc_feature    112388. .112496
                     /note="Big dye primer and dGTP big dye chemistries only."
BASE COUNT    39295 a  26385 c  24794 g  34044 t
ORIGIN


  Query Match           28.5%;  Score 493.6;  DB 9;  Length 124518;
  Best Local Similarity   47.5%;  Pred. No. 1.7e-109;
  Matches  486;  Conservative  216;  Mismatches  285;  Indels   36;  Gaps    2;

Qy    744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
          :||:|| ||::: ||:||: |::: ||:||:||:|| || ||||| :| :|  |:|| |

Db  66716 GGAAGACATTTCTAAGCAGAAAAGCATTCAAGAGGTGACTTGGGTACTGTTAGAGGCATC 66657

Qy    804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
          ::: ||:||::| ||: | ||: |:||::: :   |:||::| || || ||:||:|| ||
Db  66656 CAGTTTTATAAGGGAAACAGAGTATAAAAGTTTGAAAAATTTGCAGCCTGACAATGTGAT 66597

Qy    864 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
          : |:||:||:| :|| ||::: ||:|  ||: |:||::| || || ||:||:||:||: |
Db  66596 AGAAAAGAAAATCCCATTTTCTGAGGAGAAAGTCAAGCTGGCTGCAGAAATTTGCATAAG 66537

Qy    924 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 983
          :||:||:||::  ||:|| ||:|  ||:||:||:|| | :||: |:::  || | ||:||
Db  66536 TAACGAGGAGCCAAATGTTAATCACCAAGACAATGGAGGAAATGTCTCCAGGGCATGTCA 66477

Qy    984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naartggtt 1040
          ::| :: :: ||:|| :  |:  ::: :  || |||:  |  :| :|    ||:||||
Db  66476 GAGGTCATCACAGCATCCCCTCCCATCACAGGCCTGGAAGCTTAGGAGGAAAAAATGGTT 66417

Qy   1041 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 1100
          :  :|| |  ||  | || :: :| ||:||:|| ||:|| :| ||::| ||  | ||: |
Db  66416 TCATGGAAGGGGCACAGGGTCCCTGTGTTGTGTGCAGCCTAGGGACTTGGTGTCCTGCAT 66357

Qy   1101 nccngtnaayws----------------------------------ngcngtngcnwsnga 1127
            || |   : :                                    || || || :: ||
Db  66356 CCCAGCTGCTCCAGCTGTGGCTGAAAGGGGCCAATGTGGAGTTCAGGCTGTGGCTTCAGA 66297

Qy   1128 rggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraa 1187
          :|| || :: || ||:|| |||||::| ||    |  ||  |:|| ||  | ||  |:||
Db  66296 GGGTGCAAGCCCTAAGCCTTGGCAGCTTCCACATGATGTTAAGCCTGTGAGTGCACAGAA 66237

Qy   1188 rwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntg 1247
          ::: :| | :| :|| ||||:|| |    ::| ||:||:| :||:||:|| ||: | ||
Db  66236 ATCAAGAACTGGGGTTTGGGAACCTCTGTCTAGATTTCAGAGGATATATGGTAATACCTG 66177

Qy   1248 gatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngt 1307
          |||||| :| ||:||:||:|| ||  | |  ||    :: || :  || :: || :| |
Db  66176 GATGCCCAGACAGAAGTTTGCTGTATGGGAGGGGGTCTCATGAAAAACCTCTGCTAGGGC 66117

Qy   1308 ngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnws 1367
            ||   : |:|| ||:|| || |  ||:|   | ||::| || || :: || || ||  :
Db  66116 AGTGTGGGAGGGAAATGTGGGGTCAGAGCTGTCACACAGAGTCCCTACTGGGGCACCGCC 66057

Qy   1368 nwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacnga 1427
          ::  | || || :  :|  : ||  :: :: :| :  ||:|| || :| :: || ||
Db  66056 TAGTGGAGCTGTGAAAAGTGGGCCACTATCCTCCAGACCCCAGAATGGTAGATCCACTGA 65997

Qy   1428 ywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngc 1487
          ::: :| || ||:|  || | :||: : || ||:|| ||:||:||:|| || ||:|| ||
Db  65996 CAGCTTGCACCATGCACCTGGAAAAGCCACAGACACTCAATGCCAGCCTGTGAAAGCAGC 65937

Qy   1488 nggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnat 1547
            || | |||::: || || | :||:||    |  || ||::|  | ||:||  | || |
Db  65936 TGGGAGGGAGACTGTACCATGCAAGACCACAGGGGCAGAGCTGCCCAAGACCATGGGAAC 65877

Qy   1548 htayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttygg 1607
          : |::| :| ||:  ||:||::| ||:|| :| ||:|| || ||: | ||:||:||:||
Db  65876 CCACCTCTTGCATCAGCATGACCTGGATGTGAGACATGGAGTCAAAGGAGATCATTTTGG 65817

Qy   1608 ngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytntg 1667
            || :| :| ||:||:| :||  | || ||::  || | :|||||  |     || :| ||
Db  65816 AGCTTTAAGATTTGACTTCCCTGCTGGATTTCAGACTTGCATGGGCCTGTAGCCCCTTTG 65757

Qy   1668 yttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1727
          :||:|| ||:||:  :|| ||:|| |  |   | ||:|| ||:||::| ||::  ||:||
Db  65756 TTTTGGCCAATTTCCCCCATTTGGAATGGGTATATTTACCCAATGCCTGTACCCCCATTG 65697

Qy   1728 yat 1730
          :||
Db  65696 TAT 65694


RESULT   5
AC008799
```

```
LOCUS       AC008799              123070 bp    DNA     linear   HTG 18-JUL-2000
DEFINITION  Homo sapiens chromosome 5 clone CTD-2061E19, WORKING DRAFT
            SEQUENCE, 8 ordered pieces.
ACCESSION   AC008799
VERSION     AC008799.4  GI:9256046
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 123070)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 123070)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 18, 2000 this sequence version replaced gi:7709316.
            -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 651088
            Center clone name: CITB-H1_2061E19
            -------
            Summary Statistics
            Consensus quality: 116295 bases at least Q40
            Consensus quality: 121288 bases at least Q30
            Consensus quality: 122086 bases at least Q20
            Estimated insert size: 123000; pulse field gel estimation
            Estimated insert size: 122770; sum-of-contigs estimation
            Quality coverage:  6.04 in Q20 bases; pulse field gel estimation
            Quality coverage:  6.05 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * conslsts of 8 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1     9436: contig of 9436 bp in length
            *     9437     9536: gap of unknown length
            *     9537    26537: contig of 17001 bp in length
            *    26538    26637: gap of unknown length
            *    26638    45958: contig of 19321 bp in length
            *    45959    46058: gap of unknown length
            *    46059    68657: contig of 22599 bp in length
            *    68658    68757: gap of unknown length
            *    68758    72152: contig of 3395 bp in length
            *    72153    72252: gap of unknown length
            *    72253    81069: contig of 8817 bp in length
            *    81070    81169: gap of unknown length
            *    81170   121547: contig of 40378 bp in length
            *   121548   121647: gap of unknown length
            *   121648   123070: contig of 1423 bp in length.
FEATURES             Location/Qualifiers
     source          1. .123070
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2061E19"
                     /clone_lib="CalTech human BAC library D"
BASE COUNT    37763 a  24223 c  23855 g  36528 t    701 others
ORIGIN
```

Query Match 28.5%; Score 493; DB 2; Length 123070;

Best Local Similarity 46.7%; Pred. No. 2.3e-109;
Matches 495; Conservative 220; Mismatches 309; Indels 36; Gaps 2;

```
Qy    706 aayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcar 765
          ||: |:|  :|    :| ||  |  |:           :||:||:||::: ||:||:||:
Db 107907 AACTTTTAACTTGAAAGAAATGATTTAGGGTATCTGGCAGAAGAAATTTCTAAGCAGCAA 107966

Qy    766 wsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngar 825
          :: ||:| :||:|| || ||||| :| :| ||:|  ||::: ||:||::| ||:|| ||:
Db 107967 AGCATTCCAGAGGTGACTTGGGTGCTGTTAAAGGTATTCAGTTTTATAAGGGAAGCAGAG 108026

Qy    826 cayaarwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccnttywsn 885
          ||:||:::    ||:||::| ||  | ||:||:|| ||: |:||:||:| :|| ||:::
Db 108027 CATAAAAGTCTGGAAAATTTGCAGTCCGACAATGTGATAGAAAAGAAAATTCCATTTTCT 108086

Qy    886 garggnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaay 945
          ||:|  ||:||:||::| || || ||:|| ||:||: |:||: |:||::| ||: | ||:
Db 108087 GAGGAGAAATTCAAGCTGGCTGCAGAAATGTGCATAAGTAACAAGGAGCTGAATATTAAT 108146

Qy    946 ccncargayaaygggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnath 1005
          || ||:||:||:|| ||:||:||:::   | | ||:||::| :: :: | :||: :  |:
Db 108147 CCCCAAGACAATGGGGAAAACATCTCCAAGGCATGTCAGAGGTCTTCACGGCAGCCCCTC 108206

Qy   1006 aarwsnytngcntggmgnccnmgnmg---naartggttytgyggnacnggnccnggnwsn 1062
            ::: :   |    :| || :| :|    ||:|||||:||:||  | || || || ::
Db 108207 CCATCATAAACCCAAAGGCCTAGGAGGAAAAAGTGGTTTTGTGGGCCAGGCCCAGGGTCC 108266

Qy   1063 ytntgytgygtncarccnmgngayytngtnccntgygtnccngtnaa------------- 1109
          :| ||:||:|| ||:|  :| ||::| || || ||: | || |
Db 108267 TTGTGCTGTGTGCAGCTTAGGGACTTGGTGCCCTGCATCCCAGCTGCTCCAGCCATGGCT 108326

Qy   1110 --------------------ywsngcngtngcnwsngarggngcnwsnccnaarccntgg 1149
                              : : || || || :: ||:|| || :: || ||:|| |||
Db 108327 GAAAGGGGCCAACATAGAGCTCGGGCTGTGGCTTCAGAAGGTGCAAGCCCCAAGCCTTGG 108386

Qy   1150 carytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggar 1209
          ||::| || :: || || ||:|| |   | ||  |:|||:: :  ||:| :|| |||||:
Db 108387 CAGCTTCCATGTGGTGTTGAGCCTGCAAGTGCACAGAAGTCACAAATTGGGGTTTGGGAA 108446

Qy   1210 ccnccnathmgnttycaraarathtayggnaaycentggatgccnmgncaraarttygcn 1269
          || ||   ::| || ||:| :|| ||:|| ||: |  |||||  | :| ||:||:||:||
Db 108447 CCTCCACCTAGATTGCAGATGATGTATGGAAATGCCCGGATGACCAGGCAGAAGTTTGCT 108506

Qy   1270 gtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggn 1329
          |  || |  ||  : :: |||:| || :: || :| |  || ||:||:   ||:|| ||
Db 108507 GCAGGGGCAGGGCCTTCATGGAGAACCTCTGCTAGGGCAGTGCAGAAGAAAAATGTAGGG 108566

Qy   1330 tgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsn 1389
          |  ||:||  | ||::| || || ::  | || |   : ::  | || || :| :| ::
Db 108567 TCAGAGCCCACACACAGAGTACCCACTAGAGCACTGCCTAGTGGAGCTGTGAGAAGAAGG 108626

Qy   1390 ccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngar 1449
          || || :: :: :| :| ||:|| || :| :: ||  |::: :| ||   :|| || | :
Db 108627 CCACCATCCTCCAGACTCCAGAATGGTAGGTCCACCAACAGCTTGCACTGTGTACCTGGA 108686

Qy   1450 aarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntay 1509
          ||: : || | :|| || ||:||:|| || ||:|| ||  | | |||: :  | || | :
Db 108687 AAAGCCACAGGCACTCATTGCCAGCCTGTGAAAGCAGCCAGGAGGGAGGCTATCCCCTGC 108746

Qy   1510 aaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygay 1569
          ||:      |  ||  |::|  | ||:||  | || | :||::| :| | :   | :||:
Db 108747 AAAGTCACAGGGGCGAAGCTGTCCAAGACCTTGGGAACCTACCTCTTGCCTCAGCGTGAC 108806

Qy   1570 ytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccn 1629
          :  ||:|| :| ||:|| || ||: | ||:||:||:|| || :| :| ||:||:||:||
Db 108807 CCGGATGTGAGACATGGAGTCAAAGGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCT 108866

Qy   1630 acnggnttymgnacntayatgggnccngtnccnytntgyttyggncarttyttyccnntty 1689
           | || ||::  ||      ||| |     || :| ||:||:|| ||:||: |:|| ||:
Db 108867 GCTGGATTTCAGACTGCATGGGGCCTGTAGCCCCTTTGCTTTGGCCAATTTCTCCCATTT 108926

Qy   1690 ggnacngcngtnttyacncartgyytntayytncaytgya 1729
          || || || || ||:|| | :||::| ||::   |:||:|
```

Db 108927 GGAACAGCTGTATTTACCCCATGCCTGTACCCCGATTGTA 108966

RESULT 6
AC044889/c
LOCUS AC044889 189768 bp DNA linear HTG 22-MAY-2000
DEFINITION Homo sapiens chromosome 2 clone RP11-792C1 map 2, WORKING DRAFT SEQUENCE, 36 unordered pieces.
ACCESSION AC044889
VERSION AC044889.2 GI:8016676
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189768)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 2, clone RP11-792C1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189768)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On May 22, 2000 this sequence version replaced gi:7543856.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L9594
Center clone name: 792_C_1
---------- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 168749 bases at least Q40
Consensus quality: 180047 bases at least Q30
Consensus quality: 184014 bases at least Q20
Insert size: 190000; agarose-fp
Insert size: 186268; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1015: contig of 1015 bp in length
* 1016 1115: gap of 100 bp
* 1116 2284: contig of 1169 bp in length
* 2285 2384: gap of 100 bp
* 2385 3898: contig of 1514 bp in length
* 3899 3998: gap of 100 bp
* 3999 5259: contig of 1261 bp in length
* 5260 5359: gap of 100 bp
* 5360 6912: contig of 1553 bp in length
* 6913 7012: gap of 100 bp
* 7013 8145: contig of 1133 bp in length
* 8146 8245: gap of 100 bp
* 8246 9694: contig of 1449 bp in length
* 9695 9794: gap of 100 bp
* 9795 11004: contig of 1210 bp in length
* 11005 11104: gap of 100 bp
* 11105 12907: contig of 1803 bp in length
* 12908 13007: gap of 100 bp
* 13008 14916: contig of 1909 bp in length
* 14917 15016: gap of 100 bp
* 15017 17114: contig of 2098 bp in length
* 17115 17214: gap of 100 bp
* 17215 18527: contig of 1313 bp in length
* 18528 18627: gap of 100 bp
* 18628 20561: contig of 1934 bp in length
* 20562 20661: gap of 100 bp
* 20662 22484: contig of 1823 bp in length
* 22485 22584: gap of 100 bp
* 22585 25287: contig of 2703 bp in length
* 25288 25387: gap of 100 bp
* 25388 27603: contig of 2216 bp in length
* 27604 27703: gap of 100 bp
* 27704 30020: contig of 2317 bp in length
* 30021 30120: gap of 100 bp
* 30121 33472: contig of 3352 bp in length
* 33473 33572: gap of 100 bp
* 33573 38084: contig of 4512 bp in length
* 38085 38184: gap of 100 bp
* 38185 41069: contig of 2885 bp in length
* 41070 41169: gap of 100 bp
* 41170 45266: contig of 4097 bp in length
* 45267 45366: gap of 100 bp
* 45367 51092: contig of 5726 bp in length
* 51093 51192: gap of 100 bp
* 51193 55262: contig of 4070 bp in length
* 55263 55362: gap of 100 bp
* 55363 60982: contig of 5620 bp in length
* 60983 61082: gap of 100 bp
* 61083 67628: contig of 6546 bp in length
* 67629 67728: gap of 100 bp
* 67729 72690: contig of 4962 bp in length
* 72691 72790: gap of 100 bp
* 72791 79286: contig of 6496 bp in length
* 79287 79386: gap of 100 bp
* 79387 83451: contig of 4065 bp in length
* 83452 83551: gap of 100 bp
* 83552 92170: contig of 8619 bp in length
* 92171 92270: gap of 100 bp
* 92271 96333: contig of 4063 bp in length
* 96334 96433: gap of 100 bp
* 96434 103218: contig of 6785 bp in length
* 103219 103318: gap of 100 bp
* 103319 112553: contig of 9235 bp in length
* 112554 112653: gap of 100 bp
* 112654 123239: contig of 10586 bp in length
* 123240 123339: gap of 100 bp
* 123340 137921: contig of 14582 bp in length
* 137922 138021: gap of 100 bp
* 138022 157913: contig of 19892 bp in length
* 157914 158013: gap of 100 bp

```
          *   158014   189768: contig of 31755 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189768
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-792C1"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1015
                     /note="assembly_fragment"
     misc_feature    1116. .2284
                     /note="assembly_fragment"
     misc_feature    2385. .3898
                     /note="assembly_fragment"
     misc_feature    3999. .5259
                     /note="assembly_fragment"
     misc_feature    5360. .6912
                     /note="assembly_fragment"
     misc_feature    7013. .8145
                     /note="assembly_fragment"
     misc_feature    8246. .9694
                     /note="assembly_fragment"
     misc_feature    9795. .11004
                     /note="assembly_fragment"
     misc_feature    11105. .12907
                     /note="assembly_fragment"
     misc_feature    13008. .14916
                     /note="assembly_fragment"
     misc_feature    15017. .17114
                     /note="assembly_fragment"
     misc_feature    17215. .18527
                     /note="assembly_fragment"
     misc_feature    18628. .20561
                     /note="assembly_fragment"
     misc_feature    20662. .22484
                     /note="assembly_fragment"
     misc_feature    22585. .25287
                     /note="assembly_fragment"
     misc_feature    25388. .27603
                     /note="assembly_fragment"
     misc_feature    27704. .30020
                     /note="assembly_fragment"
     misc_feature    30121. .33472
                     /note="assembly_fragment"
     misc_feature    33573. .38084
                     /note="assembly_fragment"
     misc_feature    38185. .41069
                     /note="assembly_fragment"
     misc_feature    41170. .45266
                     /note="assembly_fragment"
     misc_feature    45367. .51092
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    51193. .55262
                     /note="assembly_fragment"
     misc_feature    55363. .60982
                     /note="assembly_fragment"

  Query Match             28.5%;  Score 493;  DB 2;  Length 189768;
  Best Local Similarity   46.7%;  Pred. No. 2.7e-109;
  Matches  495;  Conservative  220;  Mismatches  309;  Indels   36;  Gaps    2;

Qy    706 aayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcar 765
          ||: |:|  :|    :| ||  |  |:           :||:||:|||::: ||:|||:||:
Db  86595 AACTTTTAACTTGAAAGAAATGATTTAGGGTATCTGGCAGAAGAAATTTCTAAGCAGCAA 86536

Qy    766 wsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngar 825
          :: ||:| :||:|| || ||||| :| :| ||:|  ||::: ||:||::| ||:|| ||:
Db  86535 AGCATTCCAGAGGTGACTTGGGTGCTGTTAAAGGTATTCAGTTTTATAAGGGAAGCAGAG 86476

Qy    826 cayaarwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccnttywsn 885
          ||:|||:::    ||:|||::| ||  | ||:||:|| ||: |:|||:||:| :|| ||:::
Db  86475 CATAAAAGTCTGGAAAATTTGCAGTCCGACAATGTGATAGAAAAGAAAATTCCATTTTCT 86416

Qy    886 garggnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaay 945
          ||:|  ||:||:||::| || || ||:|| ||:||: |:||: |:||::| ||: | ||:
Db  86415 GAGGAGAAATTCAAGCTGGCTGCAGAAATGTGCATAAGTAACAAGGAGCTGAATATTAAT 86356

Qy    946 ccncargayaaygngaraayathwsntggacntgycarmgnwsnwsncarcarwsnath 1005
          || ||:||:||:|| ||:||:||::   | | ||:||::| :: :: | :||: :  |:
Db  86355 CCCCAAGACAATGGGGAAAACATCTCCAAGGCATGTCAGAGGTCTTCACGGCAGCCCCTC 86296

Qy   1006 aarwsnytngcntggmgnccnmgnmg---naartggttytgyggnacnggnccnggnwsn 1062
            ::: :   |    :| || :| :|    ||:|||||:||:||  | || || || ::
Db  86295 CCATCATAAACCCAAAGGCCTAGGAGGAAAAAGTGGTTTTGTGGGCCAGGCCCAGGGTCC 86236

Qy   1063 ytntgytgygtncarccnmgngayytngtnccntgygtnccngtnaa---------- 1109
          :| ||:||:|| ||:|  :| ||::| || || ||: | || |
Db  86235 TTGTGCTGTGTGCAGCTTAGGGACTTGGTGCCCTGCATCCCAGCTGCTCCAGCCATGGCT 86176

Qy   1110 -------------------ywsngcngtngcnwsngarggngcnwsnccnaarccntgg 1149
                              : : || || || :: ||:|| || :: || ||:|| |||
Db  86175 GAAAGGGGCCAACATAGAGCTCGGGCTGTGGCTTCAGAAGGTGCAAGCCCCAAGCCTTGG 86116

Qy   1150 carytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggar 1209
          ||::| || :: || || ||:|| |   | ||  |:||::: :  ||:| :|| |||||:
Db  86115 CAGCTTCCATGTGGTGTTGAGCCTGCAAGTGCACAGAAGTCACAAATTGGGGTTTGGGAA 86056

Qy   1210 ccnccnathmgnttycaraarathtayggnaayccntggatgccnmgncaraarttygcn 1269
          || ||   ::| || ||:| :|| ||:|| ||: |  |||| | :| ||:||:||:||
Db  86055 CCTCCACCTAGATTGCAGATGATGTATGGAAATGCCCGGATGACCAGGCAGAAGTTTGCT 85996

Qy   1270 gtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggn 1329
          |  || |  ||  : :: |||:| || :: || :| |  || ||:||:   ||:|| ||
Db  85995 GCAGGGGCAGGGCCTTCATGGAGAACCTCTGCTAGGGCAGTGCAGAAGAAAAATGTAGGG 85936

Qy   1330 tgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsn 1389
          |  ||:||  | ||:| || || ::  | || |   : ::  | || || :| :| ::
Db  85935 TCAGAGCCCACACACAGAGTACCCACTAGAGCACTGCCTAGTGGAGCTGTGAGAAGAAGG 85876

Qy   1390 ccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngar 1449
          || || :: :: :| :| ||:|| || :| :: ||  |::: :| ||   :|| || | :
Db  85875 CCACCATCCTCCAGACTCCAGAATGGTAGGTCCACCAACAGCTTGCACTGTGTACCTGGA 85816

Qy   1450 aarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntay 1509
          ||: : || | :|| || ||:||:|| || ||:|| ||  | | ||| : :  | || | :
Db  85815 AAAGCCACAGGCACTCATTGCCAGCCTGTGAAAGCAGCCAGGAGGGAGGCTATCCCCTGC 85756

Qy   1510 aaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygay 1569
          ||:      |  ||  |::|  | ||:||  | || | :||::| :| | :   | :||:
Db  85755 AAAGTCACAGGGGCGAAGCTGTCCAAGACCTTGGGAACCTACCTCTTGCCTCAGCGTGAC 85696

Qy   1570 ytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccn 1629
          :  ||:|| :| ||:|| || ||: | ||:||:||:|| || :| :| ||:||:||:||
Db  85695 CCGGATGTGAGACATGGAGTCAAAGGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCT 85636

Qy   1630 acnggnttymgnacntayatgggnccngtnccnytntgyttyggncarttyttyccntty 1689
           | || ||::  ||      ||| |     || :| ||:||:|| ||:||: |:|| ||:
Db  85635 GCTGGATTTCAGACTGCATGGGGCCTGTAGCCCCTTTGCTTTGGCCAATTTCTCCCATTT 85576

Qy   1690 ggnacngcngtnttyacncartgyytntayytncaytgya 1729
          || || || || ||:|| | :||::| ||::   |:||:|
Db  85575 GGAACAGCTGTATTTACCCCATGCCTGTACCCCGATTGTA 85536


RESULT   7
AC017063
LOCUS       AC017063              198545 bp    DNA     linear   HTG 02-NOV-2001
DEFINITION  Homo sapiens chromosome 4 clone RP11-354H17, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC017063
VERSION     AC017063.7  GI:16596637
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 198545)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 198545)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Nov 2, 2001 this sequence version replaced gi:15741601.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information ----------------
            Center project name: H_NH0354H17
            -----------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 35%
            Sequencing vector: plasmid; 65%
            Chemistry: Dye-primer ET; 34% of reads
            Chemistry: Dye-terminator Big Dye; 66% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 197437 bases at least Q40
            Consensus quality: 198145 bases at least Q30
            Consensus quality: 198648 bases at least Q20
            Insert size: 200000; agarose-fp
            Insert size: 200801; sum-of-contigs
            Quality coverage: 10.58 in Q20 bases; agarose-fp
            Quality coverage: 10.55 in Q20 bases; sum-of-contigs
            -----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1174: contig of 1174 bp in length
            *     1175     1274: gap of unknown length
            *     1275     2716: contig of 1442 bp in length
            *     2717     2816: gap of unknown length
            *     2817     4952: contig of 2136 bp in length
            *     4953     5052: gap of unknown length
            *     5053   198545: contig of 193493 bp in length.
FEATURES             Location/Qualifiers
     source          1. .198545
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-354H17"
     misc_feature    1. .1174
                     /note="assembly_name:Contig24"
     misc_feature    1275. .2716
                     /note="assembly_name:Contig38"
     misc_feature    2817. .4952
                     /note="assembly_name:Contig42"
     misc_feature    5053. .198545
                     /note="assembly_name:Contig44
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    62963 a  36710 c  37248 g  61324 t    300 others
ORIGIN


  Query Match           28.3%;  Score 490.4;  DB 2;  Length 198545;
  Best Local Similarity   47.0%;  Pred. No. 1.2e-108;
  Matches  498;  Conservative  213;  Mismatches  314;  Indels   34;  Gaps    2;

Qy    706 aayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcar 765
          ||: |:|  :|   |:|  |  |  |:           :||: |:||::: ||:||:||:
Db  83103 AACTTTTAACTTGGGAGAGATGATTTAGGGTATCTGGCAGAATAAATTTCTAAGCAGCAA 83162

Qy    766 wsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngar 825
          :: ||:||:|| || || ||||| :| :| ||:|| ||::: ||:||::| ||:|| ||:
Db  83163 AGCATTCAAGACGTGACTTGGGTGCTGTTAAAGGCATTCAGTTTTATAAGGGAAGCAGAG 83222

Qy    826 cayaarwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccnttywsn 885
          ||:||::  :  ||:||::| || || ||:||:|| ||: |:||:||:||:|| ||:::
Db  83223 CATAAAAATTTGGAAAATTTGCAGCCTGACAATGTGATAGAAAAGAAAAACCCATTTTCT 83282

Qy    886 garggnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaay 945
          ||:|  ||:|| ||::| || || ||:||:| :||: |:||:||:|| :| ||:|| ||:
Db  83283 GAGGAGAAATTGAAGCTGGCTGCAGAAATTTACATAAGTAATGAGGACCTGAATGTTAAT 83342

Qy    946 ccncargayaayggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnath 1005
          || ||:||:||:|| ||:||: |:::     | ||:||::| :: :  ||:||: :  |:
Db  83343 CCCCAAGACAATGGGGAAAATGTCTCCAAAGCATGTCAGAGGTCTTAACAGCAGCCCCTA 83402

Qy   1006 aarwsnytngcntggmgnccnmgnmgnaartggttytgyggnacnggnccnggnwsnytn 1065
            ::: :  || || :| || :| :| ||:||||||:| :||  | || || || :: :|
Db  83403 CCATCACAGGCCTGAAGGCCTAGGAGAAAATGGTTTTTCGGGCCAGGCCCAGGGTCCCTG 83462

Qy   1066 tgytgygtncarccnmgngayytngtnccntgygtnccngtnaayws---------- 1112
          ||:||:|| ||: | :|  |::| || || ||: | |  |  | : :
Db  83463 TGCTGTGTGCAGTCTAGGTACTTGGTGCCCTGCATCCTAGCCACTCCAGACATGACTAAA 83522

Qy   1113 --------------------ngcngtngcnwsngarggngcnwsnccnaarccntggcar 1152
                               || || || :: ||:|| |  ::    ||:|| |||| :
Db  83523 AGGGGCCAAGGTACAGCTCAGGCTGTTGCTTCAGAGGGTGGAAGCTGCAAGCCTTGGCGG 83582

Qy   1153 ytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarccn 1212
          :| || :  ||  | ||:|| |  ||  |  |:||::: :| ||: |:|| |||||:||
Db  83583 CTTCCAAATGGTATTGAACCTGCGGGTTCACAGAAGTCAAGAATTAAGGTTTGGGAACCT 83642

Qy   1213 ccnathmgnttycaraarathtayggnaaycentggatgccnmgncaraarttygcngtn 1272
          ||   ::| ||:||:||:|| ||:|| ||: | |||||||| :| ||:|| ||:|| |
Db  83643 CCGCCTAGATTTCAGAAAATGTATGGAAATGCCTGGATGCCCAGGCAGAATTTTGCTGCA 83702

Qy   1273 ggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntgg 1332
          || || ||  : :: |||:| || :: || :| |  || ||:||:|| ||:||  | |
Db  83703 GGAGTGGGGCCCTCATGGAGAACCTCTGCTAGGGCAGTGCAGAAGGGAAATGTGAGGTCA 83762

Qy   1333 garccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccn 1392
          ||:|| || ||::| |  || :: || || ||  : ::  | || || :  :|  :  |
Db  83763 GAGCCCCCACACAGAGCCCCTACTGGGGCACCACCTAGTGGAGCTGTGAAAAGAGGGACA 83822

Qy   1393 ccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraar 1452
          |  :: :: :| :  ||:|| || :| :: || || :: :| || ||:|  |  | :||:
Db  83823 CTGTCCTCCAGACCCCAGAATGGTAGATCCACTGAGAGCTTGCACCATGCACTTGGAAAA 83882

Qy   1453 wsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntayaar 1512
            : || ||:||  |:  :||: |  | ||:|| ||  | | || : : || || ||:||:
Db  83883 GCCACAGACACTGAACACCAGACCATGAAAGCAGCCAGGAGGGTGGCTGTACCCTACAAA 83942

Qy   1513 acngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygayytn 1572
            |    |  |   |::|  | ||:||    || | : |::| :| ||:   ||:||::|
Db  83943 GCCACAGGGGTGAAGTTGGCCAAGACCACAGGAACCCACCTCTTGCATCAGCATGACTTG 84002

Qy   1573 gaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacn 1632
          ||:|| :| ||:|| || ||: |  |: |:||:|| || :| :| ||:||:||:|| ||
Db  84003 GATGTGAGACATGGTGTCAAAGGACATTATTTTGGAGCTTTAAGATTTGACTGCCCCACT 84062

Qy   1633 ggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnttygg 1691
          || ||: | || | :| ||| |  ||  || :| ||:|| || ||:| : |: | ||:||
Db  84063 GGATTTTGGACTTGCACGGGGCTTGTAACCCCTTTGTTTGGGCCAATGTCTCTCATTTGG 84122

Qy   1692 nacngcngtnttyacncartgyytntayytncaytgyat 1730
               || || ||:| ||:||::| ||::  || ||: |
```

```
Db  84123 ATTTGCTGTATTTATCCAATGCCTGTACTGCCAATGTTT 84161


RESULT   8
AC025233/c
LOCUS       AC025233              161575 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-333E1, WORKING DRAFT
            SEQUENCE, 27 unordered pieces.
ACCESSION   AC025233
VERSION     AC025233.4  GI:8954317
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161575)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161575)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 7, 2000 this sequence version replaced gi:7801489.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0333E01
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 90%
            Sequencing vector: plasmid; 10%
            Chemistry: Dye-primer ET; 81% of reads
            Chemistry: Dye-terminator Big Dye; 19% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 144558 bases at least Q40
            Consensus quality: 149743 bases at least Q30
            Consensus quality: 152625 bases at least Q20
            Insert size: 176000; agarose-fp
            Insert size: 158975; sum-of-contigs
            Quality coverage: 3.48 in Q20 bases; agarose-fp
            Quality coverage: 3.75 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 27 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1141: contig of 1141 bp in length
            *     1142     1241: gap of unknown length
            *     1242     2424: contig of 1183 bp in length
            *     2425     2524: gap of unknown length
            *     2525     3961: contig of 1437 bp in length
            *     3962     4061: gap of unknown length
            *     4062     5712: contig of 1651 bp in length
            *     5713     5812: gap of unknown length
            *     5813     8019: contig of 2207 bp in length
            *     8020     8119: gap of unknown length
            *     8120    10208: contig of 2089 bp in length
            *    10209    10308: gap of unknown length
            *    10309    11734: contig of 1426 bp in length
            *    11735    11834: gap of unknown length
            *    11835    13265: contig of 1431 bp in length
            *    13266    13365: gap of unknown length
            *    13366    14425: contig of 1060 bp in length
            *    14426    14525: gap of unknown length
            *    14526    15609: contig of 1084 bp in length
            *    15610    15709: gap of unknown length
            *    15710    17529: contig of 1820 bp in length
            *    17530    17629: gap of unknown length
            *    17630    20351: contig of 2722 bp in length
            *    20352    20451: gap of unknown length
            *    20452    23753: contig of 3302 bp in length
            *    23754    23853: gap of unknown length
            *    23854    26537: contig of 2684 bp in length
            *    26538    26637: gap of unknown length
            *    26638    31339: contig of 4702 bp in length
            *    31340    31439: gap of unknown length
            *    31440    37041: contig of 5602 bp in length
            *    37042    37141: gap of unknown length
            *    37142    45268: contig of 8127 bp in length
            *    45269    45368: gap of unknown length
            *    45369    52323: contig of 6955 bp in length
            *    52324    52423: gap of unknown length
            *    52424    60239: contig of 7816 bp in length
            *    60240    60339: gap of unknown length
            *    60340    68385: contig of 8046 bp in length
            *    68386    68485: gap of unknown length
            *    68486    78834: contig of 10349 bp in length
            *    78835    78934: gap of unknown length
            *    78935    90422: contig of 11488 bp in length
            *    90423    90522: gap of unknown length
            *    90523   100763: contig of 10241 bp in length
            *   100764   100863: gap of unknown length
            *   100864   111672: contig of 10809 bp in length
            *   111673   111772: gap of unknown length
            *   111773   126855: contig of 15083 bp in length
            *   126856   126955: gap of unknown length
            *   126956   142557: contig of 15602 bp in length
            *   142558   142657: gap of unknown length
            *   142658   161575: contig of 18918 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161575
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /clone="RP11-333E1"
BASE COUNT    44288 a  35882 c  35114 g  43686 t   2605 others
ORIGIN


  Query Match             28.2%;  Score 488.8;  DB 2;  Length 161575;
  Best Local Similarity   48.5%;  Pred. No. 2.8e-108;
  Matches  485;  Conservative  202;  Mismatches  280;  Indels   34;  Gaps    2;

Qy    744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
          :||:||:||::: ||:||:||::  ||:||:|  || || ||||| :| :| || |  ||
Db 121612 GGAAGAAATTTCTAAGCAGCAAAACATTCAAGGTGTGACTTGGGTCCTGTTAAATGGATT 121553

Qy    804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
          ::: ||:|| :| ||:|| ||:||:||::: :: ||:||::| || || || ||:|| ||
Db 121552 CAGTTTTATGAGGGAAGCAGAGCATAAAAGTTCAGAAAATTTGCAGCCTGAGAATGTGAT 121493

Qy    864 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
          : |:||:||:||:|| ||::: ||:|  ||:||:|| :  || || ||:||:||:||: |
Db 121492 AGAAAAGAAAAACCCATTTTCTGAGGAGAAATTCAATCAGGCTGCAGAAATTTGCATAAG 121433

Qy    924 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 983
          :||: |: |::| ||:|| ||:|| ||:||:||:|| ||:||: |:::  || | ||:||
Db 121432 TAACAAGAAGCTGAATGTTAATCCGCAAGACAATGGGGAAAATGTCTCCAGGGCATGTCA 121373

Qy    984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 1043
          ::  :: :: ||:||: :  |:  ::: :  || |||:| || :| :| ||:|||||:||
Db 121372 GAAGTCTTCACAGCAGCCCCTCCCATCACAGGCCTGGAGGCCTAGGAGAAAATGGTTTTG 121313

Qy   1044 yggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtncc 1103
          :||  | || || || :: :  ||:||:|| ||:|| :| ||::| || |  ||: | ||
Db 121312 TGGGCCAGGCCCAGGGTCCCAATGCTGTGTCCAGCCTAGGGACTTGGTGCGCTGCATCCC 121253
```

```
Qy   1104 ngtnaaywsngcngtngcnwsngarg--------------------------------g 1130
          |    :   ||  | ||      :|                                 |
Db 121252 AGCCGCTCTAGCCATGGCTGAAAGGGGCCAATGTAGATAGAGCTCAAGCTGTGGCTTCAG 121193

Qy   1131 ngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwrs 1190
          |  :: || ||:|| |||||::| ||    || || ||: |  | ||  |  |:||:::
Db 121192 AGGGTGCCCTAAGCCTTGGCAGTTTCCACATGGTGTTGAGTCTTTAGGTACACAGAAGTC 121133

Qy   1191 nmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggat 1250
           :| ||:||:|| |||||:|  ||   : | ||:||:||:|| ||:|| ||: | |||||
Db 121132 AAGAATTGAGGTTTGGGAACTTCCACCTGGATTTCAGAAGATGTATGGAAACGCTTGGAT 121073

Qy   1251 gccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngt 1310
          ||| :| ||:||:||:|| |  || |  || :  :: |||:| || :: || :| |   |
Db 121072 GCCCAGCCAGAAGTTTGCTGCAGGGGCAGGGATCTCATGGAGAACCTCTGCTAGGGCACT 121013

Qy   1311 ncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnws 1370
          ||:||:|| ||:|| || | | |:|| ||  |: | || || ::  | || || :: ::
Db 121012 GCAGAAGGGAAATGTGGGGTCGAAGCCACCATATGGAGTCCCTACTCGGGCACCTTCTAG 120953

Qy   1371 nmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayws 1430
            | || || :| :|  : ||    :: :: :| :  ||:|| || :| :: ||  |:::
Db 120952 TGGAGCTGTGAGAAGAGGGCCAATGACCTCCAGACCCCAGAATGGTAGATCCACCTACAG 120893

Qy   1431 nytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcngg 1490
           :| || | : | || | :||: :  | ||:|| ||:  :||:|| || ||:|| ||  |
Db 120892 CTTGCACCGTATGCCTGGAAAAGCTGCAGACACTCAACACCAGCCCGTGAAAGCAGCCAG 120833

Qy   1491 natggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathta 1550
          | |||: : || || ||:||: |    |  |  ||::|  | ||:||  | || | : |
Db 120832 GAGGGAGGCTGTGCCCTACAAAGCCACAGGGGTGGAGTTGCCCAAGACCATGGGAACCCA 120773

Qy   1551 yytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngc 1610
           ::| :| ||:   ||:||::| ||:|| :| | :|| || ||: | ||:||:||:|| |
Db 120772 CCTCTTGCATCAGCATGACCTGGATGTGAGACCTGGAGTCAAAGGAGATCATTTTGGAGG 120713

Qy   1611 nytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgyt 1669
           :| :  ||:||:||:|   | || ||::| || | :||||| || ||  || :| ||:|
Db 120712 TTTAAAATTTGATTGCCTTGCTGGATTTCGGACTTGCATGGGCCCTGTAACCCCTTTGTT 120653

Qy   1670 tyggncarttyttyccnttyggnacngcngtnttyacncar 1710
          |:|| ||:||: |:|| ||:|| || || || ||:|| ||:
Db 120652 TTGGCCAATTTCTCCCATTTGGAACAGCTGTATTTACCCAG 120612


RESULT   9
AC087500
LOCUS       AC087500              163674 bp    DNA     linear   HTG 10-JAN-2002
DEFINITION  Homo sapiens chromosome 17 clone RP11-333E1 map 17, *** SEQUENCING
            IN PROGRESS ***, 11 unordered pieces.
ACCESSION   AC087500
VERSION     AC087500.4  GI:18104839
KEYWORDS    HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163674)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-333E1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163674)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 10, 2002 this sequence version replaced gi:13560412.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12015
                Center clone name: 333_E_1
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3134: contig of 3134 bp in length
            *     3135 3234: gap of      100 bp
            *     3235    27783: contig of 24549 bp in length
            *    27784 27883: gap of      100 bp
            *    27884    50218: contig of 22335 bp in length
            *    50219 50318: gap of      100 bp
            *    50319    54526: contig of 4208 bp in length
            *    54527 54626: gap of      100 bp
            *    54627    60200: contig of 5574 bp in length
            *    60201 60300: gap of      100 bp
            *    60301    70374: contig of 10074 bp in length
            *    70375 70474: gap of      100 bp
            *    70475    79596: contig of 9122 bp in length
            *    79597 79696: gap of      100 bp
            *    79697    86804: contig of 7108 bp in length
            *    86805 86904: gap of      100 bp
            *    86905   123464: contig of 36560 bp in length
            *   123465 123564: gap of      100 bp
            *   123565   160329: contig of 36765 bp in length
            *   160330 160429: gap of      100 bp
            *   160430   163674: contig of 3245 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163674
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-333E1"
                     /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    44782 a  37406 c  35496 g  44882 t   1108 others
ORIGIN


  Query Match           28.2%;  Score 488.8;  DB 2;  Length 163674;
  Best Local Similarity   48.5%;  Pred. No. 2.8e-108;
  Matches  485;  Conservative  202;  Mismatches  280;  Indels   34;  Gaps    2;

Qy    744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
          :||:||:||::: ||:||:||::  ||:||:|  || || ||||| :| :| || |  ||
```

```
Db 136131 GGAAGAAATTTCTAAGCAGCAAAACATTCAAGGTGTGACTTGGGTCCTGTTAAATGGATT 136190

Qy     804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
           ::: ||:|| :| ||:|| ||:||:||::: :: ||:||::| || || || ||:|| ||
Db 136191 CAGTTTTATGAGGGAAGCAGAGCATAAAAGTTCAGAAAATTTGCAGCCTGAGAATGTGAT 136250

Qy     864 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
           : |:||:||:||:|| ||::: ||:|  ||:||:|| :  || || ||:||:||:||: |
Db 136251 AGAAAAGAAAAACCCATTTTCTGAGGAGAAATTCAATCAGGCTGCAGAAATTTGCATAAG 136310

Qy     924 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 983
           :||: |: |::| ||:|| ||:|| ||:||:||:|| ||:||: |:::  || | ||:||
Db 136311 TAACAAGAAGCTGAATGTTAATCCGCAAGACAATGGGGAAAATGTCTCCAGGGCATGTCA 136370

Qy     984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 1043
           ::  :: :: ||:||: :  |:  ::: :  || |||:| || :| :| ||:|||||:||
Db 136371 GAAGTCTTCACAGCAGCCCCTCCCATCACAGGCCTGGAGGCCTAGGAGAAAATGGTTTTG 136430

Qy    1044 yggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtncc 1103
           :||  | || || || :: :  ||:||:|| ||:|| :| ||::| || |  ||: | ||
Db 136431 TGGGCCAGGCCCAGGGTCCCAATGCTGTGTCCAGCCTAGGGACTTGGTGCGCTGCATCCC 136490

Qy    1104 ngtnaaywsngcngtngcnwsngarg--------------------------------g 1130
            |   :  ||  | ||     :|                                   |
Db 136491 AGCCGCTCTAGCCATGGCTGAAAGGGGCCAATGTAGATAGAGCTCAAGCTGTGGCTTCAG 136550

Qy    1131 ngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarws 1190
            |  :: || ||:|| |||||:|| ||    || || ||: |  | ||  |  |:||:::
Db 136551 AGGGTGCCCTAAGCCTTGGCAGTTTCCACATGGTGTTGAGTCTTTAGGTACACAGAAGTC 136610

Qy    1191 nmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggat 1250
            :| ||:||:|| |||||:|  ||   : | ||:||:||:|| ||:|| ||: | |||||
Db 136611 AAGAATTGAGGTTTGGGAACTTCCACCTGGATTTCAGAAGATGTATGGAAACGCTTGGAT 136670

Qy    1251 gccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngt 1310
           ||| :| ||:||:||:|| |  || |  || :  :: |||:| || :: || :| |   |
Db 136671 GCCCAGCCAGAAGTTTGCTGCAGGGGCAGGGATCTCATGGAGAACCTCTGCTAGGGCACT 136730

Qy    1311 ncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnws 1370
           ||:||:|| ||:|| || | | |:|| ||  |: | || || ::  | || || :: ::
Db 136731 GCAGAAGGGAAATGTGGGGTCGAAGCCACCATATGGAGTCCCTACTCGGGCACCTTCTAG 136790

Qy    1371 nmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayws 1430
            | || || :| :|  : ||    :: :: :| :  ||:|| || :| :: ||  |:::
Db 136791 TGGAGCTGTGAGAAGAGGGCCAATGACCTCCAGACCCCAGAATGGTAGATCCACCTACAG 136850

Qy    1431 nytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcngg 1490
           :| || | : | || | :||: :  | ||:|| ||:  :||:|| || ||:|| ||  |
Db 136851 CTTGCACCGTATGCCTGGAAAAGCTGCAGACACTCAACACCAGCCCGTGAAAGCAGCCAG 136910

Qy    1491 natggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathta 1550
            | |||: : || || ||:||: |    |  |  ||:|  | ||:||  | || | : |
Db 136911 GAGGGAGGCTGTGCCCTACAAAGCCACAGGGGTGGAGTTGCCCAAGACCATGGGAACCCA 136970

Qy    1551 yytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngc 1610
           ::| :| ||:   ||:||::| ||:|| :| | :|| || ||: | ||:||:||:|| |
Db 136971 CCTCTTGCATCAGCATGACCTGGATGTGAGACCTGGAGTCAAAGGAGATCATTTTGGAGG 137030

Qy    1611 nytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgyt 1669
            :| :  ||:||:||:|   | || ||::| || | :|||| || ||  || :| ||:|
Db 137031 TTTAAAATTTGATTGCCTTGCTGGATTTCGGACTTGCATGGGCCCTGTAACCCCTTTGTT 137090

Qy    1670 tyggncarttyttyccnttyggnacngcngtnttyacncar 1710
           |:|| ||:||: |:|| ||:|| || || || ||:|| ||:
Db 137091 TTGGCCAATTTCTCCCATTTGGAACAGCTGTATTTACCCAG 137131


RESULT 10
AC015727/c
LOCUS       AC015727               175639 bp    DNA     linear   HTG 09-SEP-2000
DEFINITION  Homo sapiens chromosome 17 clone RP11-420A6 map 17, WORKING DRAFT
            SEQUENCE, 31 unordered pieces.
ACCESSION   AC015727
VERSION     AC015727.4  GI:10045179
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175639)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-420A6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175639)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 9, 2000 this sequence version replaced gi:6958075.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1423
                Center clone name: 420_A_6
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 87% of reads
                Sequencing vector: Plasmid; n/a; 13% of reads
                Chemistry: Dye-primer-amersham; 4% of reads
                Chemistry: Dye-terminator Big Dye; 96% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 161512 bases at least Q40
                Consensus quality: 167866 bases at least Q30
                Consensus quality: 170654 bases at least Q20
                Insert size: 170000; agarose-fp
                Insert size: 172639; sum-of-contigs
                Quality coverage: 4.9 in Q20 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    13947: contig of 13947 bp in length
            *    13948 14047: gap of      100 bp
            *    14048    14259: contig of 212 bp in length
            *    14260 14359: gap of      100 bp
            *    14360    15039: contig of 680 bp in length
            *    15040 15139: gap of      100 bp
            *    15140    15832: contig of 693 bp in length
            *    15833 15932: gap of      100 bp
            *    15933    16689: contig of 757 bp in length
            *    16690 16789: gap of      100 bp
            *    16790    18092: contig of 1303 bp in length
            *    18093 18192: gap of      100 bp
            *    18193    19993: contig of 1801 bp in length
```

```
          *    19994 20093: gap of      100 bp
          *    20094    22205: contig of 2112 bp in length
          *    22206 22305: gap of      100 bp
          *    22306    23915: contig of 1610 bp in length
          *    23916 24015: gap of      100 bp
          *    24016    26544: contig of 2529 bp in length
          *    26545 26644: gap of      100 bp
          *    26645    28851: contig of 2207 bp in length
          *    28852 28951: gap of      100 bp
          *    28952    30454: contig of 1503 bp in length
          *    30455 30554: gap of      100 bp
          *    30555    33015: contig of 2461 bp in length
          *    33016 33115: gap of      100 bp
          *    33116    35727: contig of 2612 bp in length
          *    35728 35827: gap of      100 bp
          *    35828    59255: contig of 23428 bp in length
          *    59256 59355: gap of      100 bp
          *    59356    62539: contig of 3184 bp in length
          *    62540 62639: gap of      100 bp
          *    62640    67930: contig of 5291 bp in length
          *    67931 68030: gap of      100 bp
          *    68031    73174: contig of 5144 bp in length
          *    73175 73274: gap of      100 bp
          *    73275    77427: contig of 4153 bp in length
          *    77428 77527: gap of      100 bp
          *    77528    81788: contig of 4261 bp in length
          *    81789 81888: gap of      100 bp
          *    81889    86379: contig of 4491 bp in length
          *    86380 86479: gap of      100 bp
          *    86480    91547: contig of 5068 bp in length
          *    91548 91647: gap of      100 bp
          *    91648    97783: contig of 6136 bp in length
          *    97784 97883: gap of      100 bp
          *    97884   105190: contig of 7307 bp in length
          *   105191 105290: gap of      100 bp
          *   105291   111855: contig of 6565 bp in length
          *   111856 111955: gap of      100 bp
          *   111956   119820: contig of 7865 bp in length
          *   119821 119920: gap of      100 bp
          *   119921   131299: contig of 11379 bp in length
          *   131300 131399: gap of      100 bp
          *   131400   139546: contig of 8147 bp in length
          *   139547 139646: gap of      100 bp
          *   139647   155538: contig of 15892 bp in length
          *   155539 155638: gap of      100 bp
          *   155639   173481: contig of 17843 bp in length
          *   173482 173581: gap of      100 bp
          *   173582   175639: contig of 2058 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175639
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-420A6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .13947
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    14048. .14259
                     /note="assembly_fragment"
     misc_feature    14360. .15039
                     /note="assembly_fragment"
     misc_feature    15140. .15832
                     /note="assembly_fragment"
     misc_feature    15933. .16689
                     /note="assembly_fragment"
     misc_feature    16790. .18092
                     /note="assembly_fragment"
     misc_feature    18193. .19993
                     /note="assembly_fragment"
     misc_feature    20094. .22205
                     /note="assembly_fragment"
     misc_feature    22306. .23915
                     /note="assembly_fragment"
     misc_feature    24016. .26544
                     /note="assembly_fragment"
     misc_feature    26645. .28851
                     /note="assembly_fragment"
     misc_feature    28952. .30454
                     /note="assembly_fragment"
     misc_feature    30555. .33015
                     /note="assembly_fragment"
     misc_feature    33116. .35727
                     /note="assembly_fragment"
     misc_feature    35828. .59255
                     /note="assembly_fragment"
     misc_feature    59356. .62539
                     /note="assembly_fragment"
     misc_feature    62640. .67930
                     /note="assembly_fragment"
     misc_feature    68031. .73174
                     /note="assembly_fragment"
     misc_feature    73275. .77427
                     /note="assembly_fragment"
     misc_feature    77528. .81788
                     /note="assembly_fragment"
     misc_feature    81889. .86379
                     /note="assembly_fragment"
     misc_feature    86480. .91547
                     /note="assembly_fragment"
     misc_feature    91648. .97783
                     /note="assembly_fragment"
     misc_feature    97884. .105190
                     /note="assembly_fragment"
     misc_feature    105291. .111855
                     /note="assembly_fragment"
     misc_feature    111956. .119820
                     /note="assembly_fragment"
     misc_feature    119921. .131299
                     /note="assembly_fragment"
     misc_feature    131400. .139546
                     /note="assembly_fragment"
     misc_feature    139647. .155538
                     /note="assembly_fragment"
     misc_feature    155639. .173481
                     /note="assembly_fragment"
     misc_feature    173582. .175639
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    48575 a  37483 c  38030 g  48544 t   3007 others

  Query Match           28.2%;  Score 488.8;  DB 2;  Length 175639;
  Best Local Similarity   48.5%;  Pred. No. 2.9e-108;
  Matches  485;  Conservative  202;  Mismatches  280;  Indels   34;  Gaps    2;

Qy     744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
           :||:||:||::: ||:||:||::  ||:||:|  || || ||||| :| :| || |  ||
Db    2929 GGAAGAAATTTCTAAGCAGCAAAACATTCAAGGTGTGACTTGGGTCCTGTTAAATGGATT 2870

Qy     804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
           ::: ||:|| :| ||:|| ||:||:||::: :: ||:||::| || || || ||:|| ||
Db    2869 CAGTTTTATGAGGGAAGCAGAGCATAAAAGTTCAGAAAATTTGCAGCCTGAGAATGTGAT 2810

Qy     864 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
           : |:||:||:||:|| ||::: ||:|  ||:||:|| :  || || ||:||:||:||: |
Db    2809 AGAAAAGAAAAACCCATTTTCTGAGGAGAAATTCAATCAGGCTGCAGAAATTTGCATAAG 2750

Qy     924 yaaygargarytnaaygtnaaycncargayaayggngaraayathwsntggacntgyca 983
           :||: |: |:|| ||:|| ||:|| ||:||:||:|| ||:||: |:::  || | ||:||
Db    2749 TAACAAGAAGCTGAATGTTAATCCACAAGACAATGGGGAAAATGTCTCCAGGGCATGTCA 2690

Qy     984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 1043
           ::  :: :: ||:||: :  |:  ::: :  || |||:| || :| :| ||:|||||:||
```

```
Db   2689 GAAGTCTTCACAGCAGCCCCTCCCATCACAGGCCTGGAGGCCTAGGAGAAAATGGTTTTG 2630

Qy   1044 yggnacnggnccnggnwsnytntgytgygtnccarccnmgngayytngtnccntgygtncc 1103
          :||  | || || || :: :  ||:||:|| ||:|| :| ||::| || |  ||: | ||
Db   2629 TGGGCCAGGCCCAGGGTCCCAATGCTGTGTCCAGCCTAGGGACTTGGTGCGCTGCATCCC 2570

Qy   1104 ngtnaaywsngcngtngcnwsngarg--------------------------------g 1130
           |    :   ||  | ||      :|                                 |
Db   2569 AGCCGCTCTAGCCATGGCTGAAAGGGGCCAATGTAGATAGAGCTCAAGCTGTGGCTTCAG 2510

Qy   1131 ngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarws 1190
           |  :: || ||:|| |||||::| ||    || || ||: |  | ||  |  |:||:::
Db   2509 AGGGTGCCCTAAGCCTTGGCAGTTTCCACATGGTGTTGAGTCTTTAGGTACACAGAAGTC 2450

Qy   1191 nmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaaycentggat 1250
          :| ||:||:|| |||||:|  ||   : | ||:||:||:|| ||:|| ||: | |||||
Db   2449 AAGAATTGAGGTTTGGGAACTTCCACCTGGATTTCAGAAGATGTATGGAAACGCTTGGAT 2390

Qy   1251 gccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngt 1310
          ||| :| ||:||:||:|| |  || |  || :  :: |||:| || :: || :| |   |
Db   2389 GCCCAGCCAGAAGTTTGCTGCAGGGGCAGGGATCTCATGGAGAACCTCTGCTAGGGCACT 2330

Qy   1311 ncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnws 1370
           ||:||:|| ||:|| || | | |:|| ||  |: | || || ::  | || || :: ::
Db   2329 GCAGAAGGGAAATGTGGGGTCGAAGCCACCATATGGAGTCCCTACTCGGGCACCTTCTAG 2270

Qy   1371 nmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayws 1430
            | || || :| :|  : ||    :: :: :| :  ||:|| || :| :: ||  |:::
Db   2269 TGGAGCTGTGAGAAGAGGGCCAATGACCTCCAGACCCCAGAATGGTAGATCCACCTACAG 2210

Qy   1431 nytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcngg 1490
           :| || | : | || | :||: :  | ||:|| ||:  :||:|| || ||:|| ||  |
Db   2209 CTTGCACCGTATGCCTGGAAAAGCTGCAGACACTCAACACCAGCCCGTGAAAGCAGCCAG 2150

Qy   1491 natggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathta 1550
           | |||: : || || ||:||: |    |  |  ||::|  | ||:||  | || | : |
Db   2149 GAGGGAGGCTGTGCCCTACAAAGCCACAGGGGTGGAGTTGCCCAAGACCATGGGAACCCA 2090

Qy   1551 yytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngc 1610
          ::| :| ||:   ||:||::| ||:|| :| | :|| || ||: | ||:||:||:|| |
Db   2089 CCTCTTGCATCAGCATGACCTGGATGTGAGACCTGGAGTCAAAGGAGATCATTTTGGAGG 2030

Qy   1611 nytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgyt 1669
           :| :  ||:||:||:|   | || ||::| || | :||||| || ||  || :| ||:|
Db   2029 TTTAAAATTTGATTGCCTTGCTGGATTTCGGACTTGCATGGGCCCTGTAACCCCTTTGTT 1970

Qy   1670 tyggncarttyttyccnttyggnacngcngtnttyacncar 1710
          |:|| ||:||: |:|| ||:|| || || || ||:|| ||:
Db   1969 TTGGCCAATTTCTCCCATTTGGAACAGCTGTATTTACCCAG 1929


RESULT  11
AL391823/c
LOCUS       AL391823              181842 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-319F11, *** SEQUENCING IN
            PROGRESS ***, 12 unordered pieces.
ACCESSION   AL391823
VERSION     AL391823.9  GI:11995113
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181842)
  AUTHORS   Mclay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 28, 2000 this sequence version replaced gi:11990058.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA319F11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 4% of reads
            Sequencing vector: plasmid; L08752; 95% of reads
            Chemistry: Dye-terminator ET-amersham; 3% of reads Chemistry:
            Dye-terminator Big Dye; 96% of reads
            Consensus quality: 174456 bases at least Q40
            Consensus quality: 177550 bases at least Q30
            Consensus quality: 179392 bases at least Q20
            Insert size: 180742; sum-of-contigs
            Insert size: 162786; 7.7% error; agarose-fp
            Quality coverage: 4.40x in Q20 bases; sum-of-contigs Quality
            coverage: 4.89x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     8896: contig of 8896 bp in length
            *     8897 8996: gap of      100 bp
            *     8997    22635: contig of 13639 bp in length
            *    22636 22735: gap of      100 bp
            *    22736    55625: contig of 32890 bp in length
            *    55626 55725: gap of      100 bp
            *    55726    65440: contig of 9715 bp in length
            *    65441 65540: gap of      100 bp
            *    65541    88588: contig of 23048 bp in length
            *    88589 88688: gap of      100 bp
            *    88689    94602: contig of 5914 bp in length
            *    94603 94702: gap of      100 bp
            *    94703   105410: contig of 10708 bp in length
            *   105411 105510: gap of      100 bp
            *   105511   107644: contig of 2134 bp in length
            *   107645 107744: gap of      100 bp
            *   107745   126586: contig of 18842 bp in length
            *   126587 126686: gap of      100 bp
            *   126687   155610: contig of 28924 bp in length
            *   155611 155710: gap of      100 bp
            *   155711   173103: contig of 17393 bp in length
            *   173104 173203: gap of      100 bp
            *   173204   181842: contig of 8639 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181842
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-319F11"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .8896
                     /note="assembly_fragment:02130
                     clone_end:SP6
                     vector_side:left"
     misc_feature    8997. .22635
                     /note="assembly_fragment:01909
                     fragment_chain:1"
     misc_feature    22736. .55625
                     /note="assembly_fragment:00961
                     fragment_chain:1"
     misc_feature    55726. .65440
                     /note="assembly_fragment:00129
                     fragment_chain:1"
     misc_feature    65541. .88588
                     /note="assembly_fragment:00944
                     fragment_chain:1"
     misc_feature    88689. .94602
```

```
                     /note="assembly_fragment:01651
                     fragment_chain:1"
     misc_feature    94703. .105410
                     /note="assembly_fragment:01169
                     fragment_chain:1"
     misc_feature    105511. .107644
                     /note="assembly_fragment:00035
                     fragment_chain:1"
     misc_feature    107745. .126586
                     /note="assembly_fragment:01445
                     fragment_chain:1"
     misc_feature    126687. .155610
                     /note="assembly_fragment:00583
                     fragment_chain:1"
     misc_feature    155711. .173103
                     /note="assembly_fragment:00742
                     fragment_chain:1"
     misc_feature    173204. .181842
                     /note="assembly_fragment:01461"
BASE COUNT    53809 a  39941 c  37882 g  49105 t   1105 others
ORIGIN
```

Query Match 28.2%; Score 488.6; DB 2; Length 181842;
Best Local Similarity 47.5%; Pred. No. 3.2e-108;
Matches 486; Conservative 212; Mismatches 290; Indels 36; Gaps 2;

```
Qy     743 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnt 802
           |:||:| : |: : ||:||:||::: ||:  :|| || || | ||| :| :| ||:|| |
Db   64835 AAGAAGGATTTCCTAAGCAGCAAAGCATTTGAGATGTGACTTTGGTGCTGTTAAAGGCAT 64776

Qy     803 tywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtna 862
           |::: ||:||::| ||:|| || ||:||::  :  ||:||::| || || ||: |:|  |
Db   64775 TCAGTTTTATAAGGGAAGCAGATCATAAAAATTTGGAAAATTTGCAGCCTGACTATGCAA 64716

Qy     863 thaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyatht 922
           |: |:||:||:||:|| ||::: | :|  ||:||:||::     || ||:||:||: |:
Db   64715 TAGAAAAGAAAAACCCATTTTCTGGGGAGAAATTCAAGCCTATTGCAGAAATTTGCCTAA 64656

Qy     923 gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 982
           |:| : |:||::  ||:|| ||:|| ||:||: |: | | :||: |::   || | ||:|
Db   64655 GTAGCAAGGAGCCTAATGTTAATCCCCAAGACCATCGGGGAAATGTCTACAGGCCATGTC 64596

Qy     983 armgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naartggt 1039
           |: | :: ::   :| : :  |:  ::: :  || || :| || :| :|    ||:||||
Db   64595 AGGGAACTTCATGGCTGCCCCTCCCATCACAAGCCTGCAGGCCCAGGAGTAAAAAGTGGT 64536

Qy    1040 tytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgyg 1099
           |:  :||  | || || || :  :| ||: |:|| ||:|  :| ||::| || || ||:|
Db   64535 TTCATGGGCCAGGCCCAGGGTTCCTGTGCCGTGTGCAGCACAGGGACTTGGTGCCCTGTG 64476

Qy    1100 tnccng----------------------------------tnaaywsngcngtngcnwsng 1126
           | || |                                  ::: || || || :: | |
Db   64475 TCCCAGCTGCTCCAGCTATGGCTGAAAGGGGCCAAGGTAGAGGTTGGGCTGTGGCTTCAG 64416

Qy    1127 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 1186
           |:|| || :: || ||:|| |||||::| ||  : |  || ||:|| || || ||  |:|
Db   64415 AGGGTGCAAGACCCAAGCCTTGGCAGCTTCCACGTGATGTTGAGCCTGTGGGTGCACAGA 64356

Qy    1187 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 1246
           |::: :| | :||:|| |||||:|| |    ::| ||:||:||:|| ||:|| ||: | |
Db   64355 AGTCAAGAACTGAGGTTTGGGAACCTCTGCCTAGATTTCAGAAGATGTATGGAAACGCCT 64296

Qy    1247 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 1306
           ||||||| :| ||:||:||:|| |  || || ||  : :: |||:| || :: || :| |
Db   64295 GGATGCCCAGGCAAAAGTTTGCTGCAGGTGTGGGACCCTCATGGAGAACCTCTGCTAGGG 64236

Qy    1307 tngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnw 1366
             || ||:||:|  ||:|| || |  ||:|| |  ||::| || || :: || || ||
Db   64235 CAGTGCAGAAGGAAAATGTGGGGTCAGAGCCCCACCACAGAGTCCCTACTGGGGCACCAC 64176

Qy    1367 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1426
           : ::  | || || :| :|  : || |  :: :: :| :  ||:|| |  :| :: || |
```

```
Db   64175 CTAGTGGAGCTGTGAGAAGGGGGCCACTGTCCTCCAGACCCCAGAATGATAGATCCACTG 64116

Qy    1427 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1486
           |::: :| || ||:   || | :||: : || ||:|| ||: |:||:||  | ||:|| |
Db   64115 ACAGCTTGCACCATTCACCTGGAAAAGCCACAGACACTCAACGCCAGCCCATGAAAGCAG 64056

Qy    1487 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1546
           |  | | |||: : || || | :||: |    |  |  ||::|  | ||:||  | || |
Db   64055 CCAGGAGGGAGGCTGTACCCTGCAAAGCCACAGGGGTGGAACTGCCCAAGACTATGGGAA 63996

Qy    1547 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1606
           :||::| :| ||:   ||:||::| ||:|| :| | :|| || ||: | ||:||:||:|
Db   63995 CCTACCTCTTGCATCAGCATGACCTGGATGTGAGACCTGGTGTCAAAGGAGATCATTTTG 63936

Qy    1607 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytnt 1666
             || :| :  ||:|||:||:|  || || ||::  || | :||||| || || || :| |
Db   63935 AAGCTTTAAAATTTGACTGCCTCACTGGATTTCAGACTTGCATGGGCCCTGTACCACTTT 63876

Qy    1667 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1726
           |:||:|| ||:||: |:|| ||:|| |  ||  | ||:|| ||:| ::| ||::  ||:|
Db   63875 GTTTTGGCCAATTTCTCCCATTTGGGAGGGCTATATTTACCCAGTACTTGTACCCCCATT 63816

Qy    1727 gyat 1730
           |:||
Db   63815 GTAT 63812
```

```
RESULT  12
AL353634
LOCUS       AL353634               77405 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome X clone RP11-435A2 map q21.33-22.3, ***
            SEQUENCING IN PROGRESS***, in ordered pieces.
ACCESSION   AL353634
VERSION     AL353634.4  GI:9863649
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 77405)
  AUTHORS   Mclay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 21, 2000 this sequence version replaced gi:9213111.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA435A2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 13% of reads Chemistry:
            Dye-terminator Big Dye; 86% of reads
            Consensus quality: 76604 bases at least Q40
            Consensus quality: 77098 bases at least Q30
            Consensus quality: 77320 bases at least Q20
            Insert size: 77405; sum-of-contigs
            Insert size: 85298; 1.5% error; agarose-fp
            Quality coverage: 4.54x in Q20 bases; sum-of-contigs Quality
            coverage: 4.30x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .77405
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q21.33-22.3"
                     /clone="RP11-435A2"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .77405
                     /note="assembly_fragment:00554"
BASE COUNT    25072 a  14551 c  15133 g  22649 t
ORIGIN
```

Query Match 28.2%; Score 488.2; DB 2; Length 77405;
Best Local Similarity 47.0%; Pred. No. 3e-108;
Matches 508; Conservative 225; Mismatches 310; Indels 37; Gaps 4;

```
Qy     688 ytncaycaytgyaargtnaayathtggytnatgmgnaarcarathytngcnaayaargar 747
Db    8637 CTGCCCTAGAGATCTGTGGAATTTGACCTTGAGAGAGATGATTTAGGGTATCAGGTGGAA 8696

Qy     748 garathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnttywsn 807
Db    8697 GAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGTGTCCTGTTAAAGGCACTCAGT 8756

Qy     808 ttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnathaar 867
Db    8757 TTTGCAAGGGAAGCAGAACATAAAAGTTCAGAAAATTTTCAGCCTGACAATGTGATAGAA 8816

Qy     868 aaraaraaycc-nttywsngarggnaarttyaarytngcngcngarathtgyathtgyaa 926
Db    8817 AAGAAAACCCCATTTTTCTGATGATAAATTGAAGCCAGCTGCAGAAATTTGCATAAGTAA 8876

Qy     927 ygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycarmg 986
Db    8877 CGAGGAGCCAAATGTTAATCACTAAGACCATGGGGAAAATATCTCCATGGCTTGTCAGAG 8936

Qy     987 nwsnwsncarcarwsnathaarwsnytng--cntggmgnccnmgnmgnaartggttytgy 1044
Db    8937 GTCTTCACGGCAGCCCCTCACATCACTGGCCCCCAGAGGCCTAGGAGAAAATGGTTTTAT 8996

Qy    1045 ggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtnccn 1104
Db    8997 GGGTCGGGCCCAGGGTCCCCATGCTATGTGCAGTCTAGGGACTTGGTGCCCCGTGTCCCA 9056

Qy    1105 gtnaa---------------------------------ywsngcngtngcnwsngarggn 1131
Db    9057 GCCACTCCAGCCATAACTGAAAGGGGCCAAGGTATAGCTTGGGCTGTTGCTTCAAAGGAT 9116

Qy    1132 gcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsn 1191
Db    9117 GGAAGCCCCAGGCCTTGGCAGCTTCCATGTGGTGTTGAGCCTGCAGGTGCACTGAAGTCA 9176

Qy    1192 mgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggatg 1251
Db    9177 AGAATTGAGGTTTGGGAACCTCCACCTAGATTTCAGAAGATGTATGGAAATGCCTGGATG 9236

Qy    1252 ccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtn 1311
Db    9237 CCCAGGCAAAAGTTTGCTGCAGGGGCGAGGACCTCATGGAGAACCTCTGTTAGGGTGGTG 9296

Qy    1312 caraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsn 1371
Db    9297 TGGAAGGGAAAAGTGGGGTCATAGCCGCCACACAGAGTCCCTACCAGGGGACCACCTAGT 9356

Qy    1372 mgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsn 1431
Db    9357 GGAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAATGGTAGATCCACCGACAGC 9416

Qy    1432 ytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggn 1491
Db    9417 CTGCACCGTGTGCCTGGAAAAGCCGCAGACACTCAATGCCAGCCTGTGAAAGCAGCAGGG 9476

Qy    1492 atggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtay 1551
```

```
Db    9477 AGGGAGGCTGTACCCTGCAAAGCCACAGGGGCAGAGCTGCCCAAGATCATGGGAACCAAT 9536

Qy    1552 ytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcn 1611
Db    9537 CTCTTGCATCAGCATGACCTGGATATGAGACATGGAGTCAAAGGAAATCATTTTGGAGCT 9596

Qy    1612 ytnmgnttygaytgyccnacnggnttymgnacntayatgggnccng9t-nccnytntgytt 1670
Db    9597 TTAAGATTTGACTGTCCCACTGGATCTCAGACTTCCATGGGGCCTGTAGCCTCTTTGTTT 9656

Qy    1671 yggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytgyat 1730
Db    9657 TGGCCAATATCTCCCATTTGGAATGGCTGTATTTATCCAATGCCTGTACCCCCATTGTAT 9716
```

```
RESULT  13
HSJ633H17/c
LOCUS       HSJ633H17             124531 bp    DNA     linear   PRI 07-FEB-2000
DEFINITION  Human DNA sequence from clone 633H17 on chromosome 1p31.2-32.2.
            Contains a pseudogene similar to part of MTCO1 (Cytochrome C
            oxidase 1), MTCO2 (cytochrome c oxidase II), MTATP8 (ATP synthase
            8) and GOT2 glutamic-oxaloacetic transaminase 2, mitochondrial
            (aspartate aminotransferase 2) pseudogenes, ESTs and GSSs, complete
            sequence.
ACCESSION   AL049710
VERSION     AL049710.18  GI:5650682
KEYWORDS    HTG; aspartate aminotransferase; ATP synthase; Cytochrome C
            oxidase; glutamic-oxaloacetic transaminase; GOT2; MTATP8; MTCO1;
            MTCO2.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 124531)
  AUTHORS   Moore,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 29, 1999 this sequence version replaced gi:5566565.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence is the entire insert of clone 633H17. This sequence
            has been finished according to sequence map criteria as follows. An
            attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 1, constructed by the Sanger Centre Chromosome 1
            Mapping Group. Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            633H17 is from the library RPCI-4 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2 The
            following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep.
FEATURES              Location/Qualifiers
     source           1. .124531
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="1"
                      /map="p31.2-32.2"
                      /clone_lib="RPCI-4"
                      /clone="633H17"
```

```
repeat_region   1. .234
                /note="AluSx repeat: matches 59. .292 of consensus"
misc_feature    complement(235. .546)
                /note="match: GSS: Em:AQ055811"
repeat_region   559. .760
                /note="MIR repeat: matches 74. .262 of consensus"
repeat_region   1158. .2018
                /note="L1MA4 repeat: matches 5282. .6145 of consensus"
repeat_region   2026. .2590
                /note="HERVL repeat: matches 4676. .5241 of consensus"
repeat_region   2879. .9441
                /note="L1PA2 repeat: matches 1. .6146 of consensus"
repeat_region   10000. .10124
                /note="MIR repeat: matches 34. .154 of consensus"
repeat_region   10623. .10689
                /note="MIR repeat: matches 36. .102 of consensus"
repeat_region   10690. .11014
                /note="L2 repeat: matches 2389. .2710 of consensus"
repeat_region   11136. .11262
                /note="L2 repeat: matches 2349. .2473 of consensus"
repeat_region   13420. .13522
                /note="MIR repeat: matches 114. .219 of consensus"
repeat_region   13602. .13725
                /note="MIR repeat: matches 130. .262 of consensus"
repeat_region   13739. .13987
                /note="L2 repeat: matches 2397. .2666 of consensus"
repeat_region   14368. .14486
                /note="L2 repeat: matches 2585. .2705 of consensus"
repeat_region   14656. .14764
                /note="MIR repeat: matches 124. .252 of consensus"
misc_feature    14765. .14965
                /note="match: GSS: Em:B30406"
repeat_region   14966. .15150
                /note="MIR repeat: matches 13. .217 of consensus"
repeat_region   15157. .15818
                /note="MER39b repeat: matches 14. .579 of consensus"
misc_feature    16403. .16945
                /note="match: GSS: Em:AQ454645"
misc_feature    16410. .16933
                /note="match: GSS: Em:AQ357205"
misc_feature    16410. .17037
                /note="match: GSS: Em:AQ309712"
misc_feature    complement(16459. .16861)
                /note="match: GSS: Em:AQ197233"
repeat_region   16531. .16576
                /note="23 copies 2 mer tt 74% conserved"
repeat_region   18053. .18154
                /note="MIR repeat: matches 7. .107 of consensus"
repeat_region   18377. .18492
                /note="MER39b repeat: matches 476. .574 of consensus"
misc_feature    18525. .18954
                /note="match: GSS: Em:AQ210513"
repeat_region   19026. .19126
                /note="MIR repeat: matches 10. .114 of consensus"
repeat_region   19866. .20007
                /note="MIR repeat: matches 28. .169 of consensus"
repeat_region   20273. .20334
                /note="LTR16C repeat: matches 90. .152 of consensus"
repeat_region   20403. .20505
                /note="MER81 repeat: matches 1. .110 of consensus"
repeat_region   20508. .20690
                /note="LTR16C repeat: matches 190. .376 of consensus"
repeat_region   20692. .20725
                /note="L2 repeat: matches 2660. .2693 of consensus"
repeat_region   20828. .20863
                /note="18 copies 2 mer ac 94% conserved"
repeat_region   21172. .21304
                /note="MER5A repeat: matches 40. .167 of consensus"
repeat_region   21201. .21331
                /note="MER5A repeat: matches 56. .188 of consensus"
repeat_region   21466. .21528
                /note="MER54A repeat: matches 214. .285 of consensus"
repeat_region   21952. .22077
                /note="L2 repeat: matches 2578. .2707 of consensus"
repeat_region   22083. .22259
                /note="FRAM repeat: matches -2. .175 of consensus"
repeat_region   22280. .22393
                /note="L2 repeat: matches 2541. .2645 of consensus"
repeat_region   22394. .22748
                /note="MLT1B repeat: matches 1. .390 of consensus"
repeat_region   22749. .22846
                /note="L2 repeat: matches 2645. .2745 of consensus"
repeat_region   22864. .23010
                /note="L2 repeat: matches 2329. .2429 of consensus"
repeat_region   23011. .23310
                /note="AluJb repeat: matches 1. .299 of consensus"
repeat_region   23311. .23391
                /note="L2 repeat: matches 2429. .2521 of consensus"
repeat_region   23433. .23730
                /note="MSTD repeat: matches 66. .394 of consensus"
repeat_region   23731. .25397
                /note="L1PB1 repeat: matches 4442. .6123 of consensus"
repeat_region   25477. .25556
                /note="MSTD repeat: matches 1. .85 of consensus"
repeat_region   25557. .25651
                /note="L2 repeat: matches 2585. .2692 of consensus"
repeat_region   25745. .25785
                /note="L2 repeat: matches 2702. .2742 of consensus"
repeat_region   26898. .27373
                /note="MER31A repeat: matches 1. .661 of consensus"
repeat_region   27378. .27483
                /note="MER93 repeat: matches 292. .397 of consensus"
misc_feature    27484. .27868
                /note="match: GSS: Em:AQ127976"
repeat_region   27537. .27688
                /note="MER93 repeat: matches 1. .188 of consensus"
repeat_region   27912. .28051
                /note="L1PA13 repeat: matches 6011. .6152 of consensus"
repeat_region   29900. .30121
                /note="HERVL repeat: matches 3020. .3235 of consensus"
repeat_region   30122. .30487
                /note="MLT1A2 repeat: matches 1. .374 of consensus"
repeat_region   30488. .30578
                /note="HERVL repeat: matches 2933. .3020 of consensus"
repeat_region   30579. .30796
                /note="MER58A repeat: matches 1. .221 of consensus"
repeat_region   30797. .31887
                /note="HERVL repeat: matches 1792. .2933 of consensus"
misc_feature    complement(31888. .32090)
                /note="match: GSS: Em:B58153; match: GSS: Em:B66544;
                match: GSS: Em:AQ759038"
misc_feature    32922. .33453
                /note="match: GSS: Em:AQ751315"
repeat_region   33175. .33412
                /note="L2 repeat: matches 2473. .2708 of consensus"
repeat_region   33454. .33885
                /note="MLT1C repeat: matches 6. .466 of consensus"
repeat_region   33896. .33946
                /note="MIR repeat: matches 162. .215 of consensus"
repeat_region   33947. .34302
                /note="THE1B repeat: matches 1. .364 of consensus"
repeat_region   34303. .34334
                /note="MIR repeat: matches 129. .162 of consensus"
repeat_region   34335. .34703
                /note="THE1C repeat: matches 1. .369 of consensus"
repeat_region   34704. .34830
                /note="MIR repeat: matches 20. .129 of consensus"
repeat_region   34831. .35552
                /note="L1MC2 repeat: matches 5573. .6325 of consensus"
repeat_region   36612. .36657
                /note="L2 repeat: matches 2675. .2723 of consensus"
repeat_region   36936. .37004
                /note="L2 repeat: matches 2638. .2707 of consensus"
repeat_region   37383. .37648

Query Match          28.2%;  Score 488.2;  DB 9;  Length 124531;
```

```
  Best Local Similarity   47.4%;  Pred. No. 3.5e-108;
  Matches  485;  Conservative  213;  Mismatches  290;  Indels   36;  Gaps    2;

Qy    743 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnt 802
          |:||:| : |: : ||:||:||::: ||:  :|| || || | ||| :| :| ||:|| |
Db  96092 AAGAAGGATTTCCTAAGCAGCAAAGCATTTGAGATGTGACTTTGGTGCTGTTAAAGGCAT 96033

Qy    803 tywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtna 862
          |::: ||:||::| ||:|| ||:||:|||::  :  ||:||::| || || ||: |:|  |
Db  96032 TCAGTTTTATAAGGGAAGCAGAGCATAAAAATTTGGAAAATTTGCAGCCTGACTATGCAA 95973

Qy    863 thaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyatht 922
          |: |:||:||:||:|| ||::: | :|  ||: |:||::     || ||:||:||: |:
Db  95972 TAGAAAAGAAAAACCCATTTTCTGGGGAGAAACTCAAGCCTATTGCAGAAATTTGCCTAA 95913

Qy    923 gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 982
          |:| : |:||::  ||:|| ||:|| ||:||: |: | | :||: |::   || | ||:|
Db  95912 GTAGCAAGGAGCCTAATGTTAATCCCCAAGACCATCGGGGAAATGTCTACAGGCCATGTC 95853

Qy    983 armgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naartggt 1039
          |: | :: ::   :| : :  |:  ::: :  || || :| || :| :|    ||:||||
Db  95852 AGGGAACTTCATGGCTGCCCCTCCCATCACAAGCCTGCAGGCCCAGGAGTAAAAAGTGGT 95793

Qy   1040 tytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgyg 1099
          |:  :||  | || || || :  :| ||: |:|| ||:|  :| ||::| || || ||:|
Db  95792 TTCATGGGCCAGGCCCAGGGTTCCTGTGCCGTGTGCAGCACAGGGACTTGGTGCCCTGTG 95733

Qy   1100 tnccng--------------------------------tnaaywsngcngtngcnwsng 1126
          | || |                                ::: || || || :: |
Db  95732 TCCCAGCTGCTCCAGCTATGGCTGAAAGGGGCCAAGGTAGAGGTTGGGCTGTGGCTTCAG 95673

Qy   1127 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 1186
          |:|| || :: || ||:|| |||||::| ||  : |  || ||:|| || || ||  |:|
Db  95672 AGGGTGCAAGACCCAAGCCTTGGCAGCTTCCACGTGATGTTGAGCCTGTGGGTGCACAGA 95613

Qy   1187 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 1246
          |::: :| | :||:|| |||||:|| |    ::| ||:||:||:|| ||:|| ||: | |
Db  95612 AGTCAAGAACTGAGGTTTGGGAACCTCTGCCTAGATTTCAGAAGATGTATGGAAACGCCT 95553

Qy   1247 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 1306
          ||||||| :| ||:||:||:|| |  || || ||  : :: |||:| || :: || :| |
Db  95552 GGATGCCCAGGCAAAAGTTTGCTGCAGGTGTGGGACCCTCATGGAGAACCTCTGCTAGGG 95493

Qy   1307 tngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnw 1366
            || ||:||:|  ||:|| || |  ||:|| |  ||::| || || :: || || ||
Db  95492 CAGTGCAGAAGGAAAATGTGGGGTCAGAGCCCCACCACAGAGTCCCTACTGGGGCACCAC 95433

Qy   1367 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1426
          : ::  | || || :| :|  : || |  :: :: :| :  ||:|| |  :| :: || |
Db  95432 CTAGTGGAGCTGTGAGAAGGGGGCCACTGTCCTCCAGACCCCAGAATGATAGATCCACTG 95373

Qy   1427 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1486
          |::: :| || ||:   || | :||: : || ||:|| ||: |:||:||  | ||:|| |
Db  95372 ACAGCTTGCACCATTCACCTGGAAAAGCCACAGACACTCAACGCCAGCCCATGAAAGCAG 95313

Qy   1487 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1546
          |  | | |||: : || || | :||: |    |  |  ||::|  | ||:||  | || |
Db  95312 CCAGGAGGGAGGCTGTACCCTGCAAAGCCACAGGGGTGGAACTGCCCAAGACTATGGGAA 95253

Qy   1547 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1606
           :||::| :| ||:   ||:||::| ||:|| :| | :|| || ||: | ||:||:||:|
Db  95252 CCTACCTCTTGCATCAGCATGACCTGGATGTGAGACCTGGTGTCAAAGGAGATCATTTTG 95193

Qy   1607 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytnt 1666
            || :| :  ||:||:||:|  || || ||::  || | :||||| || || || :| |
Db  95192 AAGCTTTAAAATTTGACTGCCTCACTGGATTTCAGACTTGCATGGGCCCTGTACCACTTT 95133

Qy   1667 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1726
          |:||:|| ||:||: |:|| ||:|| |  ||  | ||:|| ||:| ::| ||::  ||:|
Db  95132 GTTTTGGCCAATTTCTCCCATTTGGGAGGGCTATATTTACCCAATACTTGTACCCCCATT 95073

Qy   1727 gyat 1730
          |:||
```

```
Db  95072 GTAT 95069


RESULT  14
HSU212C1
LOCUS       HSU212C1                40714 bp    DNA     linear   PRI 23-NOV-1999
DEFINITION  Human DNA sequence from cosmid U212C1, between markers DXS366 and
            DXS87 on chromosome X.
ACCESSION   Z69722
VERSION     Z69722.1  GI:1204119
KEYWORDS    X.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 40714)
  AUTHORS   Odell,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1995) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, UK. E-mail enquires: humpub@sanger.ac.uk
COMMENT     IMPORTANT:  This sequence is the entire insert of clone U212C1. The
            true left end of clone U212C1 is at 1 in this sequence. The true
            right end of clone U212C1 is at 40714.
            U212C1 is from the human chromosome X-specific cosmid library.
FEATURES             Location/Qualifiers
     source          1. .40714
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="X"
                     /clone="LL0XNC01-212C1"
                     /clone_lib="LL0XNC01"
     repeat_region   10. .233
                     /note="L1 element fragment"
     repeat_region   388. .758
                     /note="L1 element fragment"
     repeat_region   932. .1108
                     /note="L1 element fragment"
     repeat_region   2571. .2617
                     /note="MER43 element fragment"
     repeat_region   2613. .5203
                     /note="L1 element fragment"
     repeat_region   5204. .5229
                     /note="13 copies of 2 mer 100 % conserved"
     repeat_region   5230. .5427
                     /note="MER43 element fragment"
     repeat_region   8954. .9088
                     /note="L1 element fragment"
     repeat_region   12171. .13985
                     /note="L1 element fragment"
     repeat_region   14179. .14612
                     /note="L1 element fragment"
     repeat_region   14662. .14780
                     /note="L1 element fragment"
     repeat_region   14824. .15044
                     /note="L1 element fragment"
     repeat_region   15234. .15369
                     /note="L1 element fragment"
     repeat_region   15421. .15849
                     /note="L1 element fragment"
     repeat_region   16026. .16155
                     /note="MSTA element fragment"
     repeat_region   16073. .16149
                     /note="THE1B element fragment"
     repeat_region   16204. .16495
                     /note="Alu repeat: matches 1. .308 of consensus"
     repeat_region   16532. .16609
                     /note="THE1B element fragment"
     repeat_region   16624. .16722
                     /note="MSTC element fragment"
     repeat_region   16626. .16737
                     /note="MSTA element fragment"
     repeat_region   16651. .16712
```

```
                     /note="THE1B element fragment"
     repeat_region   17060. .17653
                     /note="L1 element fragment"
     repeat_region   19720. .19797
                     /note="MER21B element fragment"
     repeat_region   19877. .19915
                     /note="MER21B element fragment"
     repeat_region   20007. .20256
                     /note="MER21B element fragment"
     repeat_region   20518. .20734
                     /note="MER30 element fragment"
     repeat_region   22661. .22730
                     /note="MLT1B element fragment"
     repeat_region   22740. .22814
                     /note="MLT1A element fragment"
     repeat_region   22917. .22974
                     /note="MLT1A element fragment"
     repeat_region   23342. .23501
                     /note="THE1B element fragment"
     repeat_region   23550. .23697
                     /note="THE1B element fragment"
     repeat_region   23725. .24279
                     /note="THR element fragment"
     repeat_region   24395. .24593
                     /note="THR element fragment"
     repeat_region   24643. .25270
                     /note="THR element fragment"
     repeat_region   25277. .25463
                     /note="THE1B element fragment"
     repeat_region   25528. .25627
                     /note="THE1B element fragment"
     repeat_region   26081. .26196
                     /note="MLT1A element fragment"
     repeat_region   26265. .26661
                     /note="LTR2 element fragment"
     repeat_region   26797. .26898
                     /note="MSTC element fragment"
     repeat_region   27949. .28241
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   28319. .28608
                     /partial
                     /note="Alu repeat: matches 308. .1 of consensus"
     repeat_region   30032. .30159
                     /note="MER5 element fragment"
     repeat_region   31794. .31980
                     /partial
                     /note="Alu repeat: matches 308. .109 of consensus"
     repeat_region   31981. .32073
                     /partial
                     /note="Alu repeat: matches 93. .1 of consensus"
     repeat_region   32232. .32735
                     /note="MER9 element fragment"
     repeat_region   35168. .35654
                     /note="MER4B element fragment"
     repeat_region   40640. .40693
                     /note="27 copies of 2 mer 94 % conserved"
BASE COUNT    11544 a   7879 c   8141 g  13150 t
ORIGIN


  Query Match           28.2%;  Score 487.6;  DB 9;  Length 40714;
  Best Local Similarity   47.0%;  Pred. No. 3.3e-108;
  Matches  507;  Conservative  225;  Mismatches  311;  Indels   36;  Gaps    4;

Qy    688 ytncaycaytgyaargtnaayathtggytnatgmgnaarcarathytngcnaayaargar 747
          :| | : |  |    ||  |  |:   :|   |:|  |  |  |:        |   :||:
Db  24003 CTGCCCTAGAGATCTGTGGAATTTGACCTTGAGAGAGATGATTTAGGGTATCAGGTGGAA 24062

Qy    748 garathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnttywsn 807
          ||:||::: ||:||:||::: ||:||:||:|| || || || :| :| ||:||  |:::
Db  24063 GAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGTGTCCTGTTAAAGGCACTCAGT 24122

Qy    808 ttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnathaar 867
          ||:  ::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||: |:
Db  24123 TTTGCAAGGGAAGCAGAACATAAAAGTTCAGAAAATTTTCAGCCTGACAATGTGATAGAA 24182

Qy    868 aaraaraaycc-nttywsngarggnaarttyaarytngcngcngarathtgyathtgyaa 926
          ||:||:| :|| ||::: || |  ||:|| ||::  || || ||:||:||:||: |:||
Db  24183 AAGAAAACCCCATTTTTCTGATGATAAATTGAAGCCAGCTGCAGAAATTTGCATAAGTAA 24242

Qy    927 ygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycarmg 986
          :||:|||::  ||:|| ||:|   |:||: |:|| ||:||:||:::   | | ||:||::|
Db  24243 CGAGGAGCCAAATGTTAATCACTAAGACCATGGGGAAAATATCTCCATGGCTTGTCAGAG 24302

Qy    987 nwsnwsncarcarwsnathaarwsnyt-ngcntggmgnccnmgnmgnaartggttytgyg 1045
          :: :: | :||: :  |:| ::: :|   |   |:| || :| :| ||:|||||:| :|
Db  24303 GTCTTCACGGCAGCCCCTCACATCACTGCCCCCAGAGGCCTAGGAGAAAATGGTTTTATG 24362

Qy   1046 gnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtnccng 1105
          |  | || || || :: :  ||:| :|| ||: | :| ||::| || ||  |:|| || |
Db  24363 GGTCGGGCCCAGGGTCCCCATGCTATGTGCAGTCTAGGGACTTGGTGCCCCGTGTCCCAG 24422

Qy   1106 tnaa--------------------------------ywsngcngtngcnwsngarggng 1132
            |                                  ::: || || || ::  |:|  |
Db  24423 CCACTCCAGCCATAACTGAAAGGGGCCAAGGTATAGCTTGGGCTGTTGCTTCAAAGGATG 24482

Qy   1133 cnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnm 1192
            :: || | :|| ||||||:| || :: || || || ||:|| |  || ||   :||::: :
Db  24483 GAAGCCCCAGGCCTTGGCAGCTTCCATGTGGTGTTGAGCCTGCAGGTGCACTGAAGTCAA 24542

Qy   1193 gnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggatgc 1252
          | ||:||:|| ||||:|| ||   ::| ||:||:||:|| ||:|| ||: | |||||||
Db  24543 GAATTGAGGTTTGGGAACCTCCACCTAGATTTCAGAAGATGTATGGAAATGCCTGGATGC 24602

Qy   1253 cnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtnc 1312
          | :| ||:||:||:|| |  || |   | :: :: |||:| || :: |  :| || ||
Db  24603 CCAGGCAAAAGTTTGCTGCAGGGGCGAGGACCTCATGGAGAACCTCTGTTAGGGTGGTGT 24662

Qy   1313 araarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnm 1372
            :||:|| || || || |   |:|| || ||::| || || ::  | |  ||  : ::
Db  24663 GGAAGGGAAAAGTGGGGTCATAGCCGCCACACAGAGTCCCTACCAGGGGACCACCTAGTG 24722

Qy   1373 gngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsny 1432
          | || || :| :|  : || || :: :: :| :  ||:|| || :| :: || ||::: :
Db  24723 GAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAATGGTAGATCCACCGACAGCC 24782

Qy   1433 tncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggna 1492
          | || | :|| || | :||: :  | ||:|| ||:||:||:|| || ||:|| || || |
Db  24783 TGCACCGTGTGCCTGGAAAAGCCGCAGACACTCAATGCCAGCCTGTGAAAGCAGCAGGGA 24842

Qy   1493 tggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtayy 1552
            |||: : || || | :||: |    |  || ||::|  | ||:|   | || | : |::
Db  24843 GGGAGGCTGTACCCTGCAAAGCCACAGGGGCAGAGCTGCCCAAGATCATGGGAACCAATC 24902

Qy   1553 tnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcny 1612
          | :| ||:   ||:||::| ||: | :| ||:|| || ||: |  |:||:||:|| || :
Db  24903 TCTTGCATCAGCATGACCTGGATATGAGACATGGAGTCAAAGGAAATCATTTTGGAGCTT 24962

Qy   1613 tnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgytty 1671
          | :| ||:||:||:|| || || | ::  || | :|||| || ||  || :| ||:||:
Db  24963 TAAGATTTGACTGTCCCACTGGATCTCAGACTTCCATGGGGCCTGTAGCCTCTTTGTTTT 25022

Qy   1672 ggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytgyat 1730
          || ||:| : |:|| ||:|| |  || || ||:|  ||:||::| ||::  ||:||:||
Db  25023 GGCCAATATCTCCCATTTGGAATGGCTGTATTTATCCAATGCCTGTACCCCCATTGTAT 25081


RESULT  15
AP004219/c
LOCUS       AP004219              159712 bp    DNA     linear   PRI 26-DEC-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 8q23, clone: KB1222D11,
            complete sequence.
ACCESSION   AP004219
VERSION     AP004219.2  GI:18146615
KEYWORDS    .
```

```
SOURCE      Homo sapiens pre-pro-B cell cell_line:FLEB 14 - 14 DNA,
            clone_lib:Keio BAC library clone:KBl222Dll.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159712)
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2001) Nobuyoshi Shimizu, Keio University, School
            of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
            160-8582, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp,
            Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT     On Jan 14, 2002 this sequence version replaced gi:15824050.
FEATURES             Location/Qualifiers
     source          1. .159712
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23"
                     /clone="KBl222Dll"
                     /cell_line="FLEB 14 - 14"
                     /cell_type="pre-pro-B cell"
                     /clone_lib="Keio BAC library"
     repeat_region   1. .4242
                     /rpt_family="LlPA2"
                     /evidence=not_experimental
     repeat_region   4668. .4733
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   5899. .5921
                     /rpt_family="(TGGA)n"
                     /evidence=not_experimental
     repeat_region   complement(6205. .6406)
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   6410. .6589
                     /rpt_family="FAM"
                     /evidence=not_experimental
     repeat_region   6619. .6648
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(6649. .6776)
                     /rpt_family="AluJo/FRAM"
                     /evidence=not_experimental
     repeat_region   6860. .7158
                     /rpt_family="AluJo"
                     /evidence=not_experimental
     repeat_region   9083. .9127
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   9188. .9226
                     /rpt_family="(TA)n"
                     /evidence=not_experimental
     repeat_region   complement(9227. .9522)
                     /rpt_family="AluSq"
                     /evidence=not_experimental
     repeat_region   9836. .9933
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   10728. .10762
                     /rpt_family="LlMAl"
                     /evidence=not_experimental
     repeat_region   10763. .11062
                     /rpt_family="AluSg"
                     /evidence=not_experimental
     repeat_region   11063. .11300
                     /rpt_family="LlMAl"
                     /evidence=not_experimental
     repeat_region   11301. .11324
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   11520. .11545
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   12507. .12560
                     /rpt_family="(TAAAAA)n"
                     /evidence=not_experimental
     repeat_region   12793. .12825
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   15792. .15852
                     /rpt_family="AluJ/FRAM"
                     /evidence=not_experimental
     repeat_region   15993. .16013
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(16014. .16312)
                     /rpt_family="AluSq"
                     /evidence=not_experimental
     repeat_region   17126. .17436
                     /rpt_family="AluSgl"
                     /evidence=not_experimental
     repeat_region   complement(17798. .17960)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   18007. .18197
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   18198. .18489
                     /rpt_family="AluSc"
                     /evidence=not_experimental
     repeat_region   18490. .18715
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   complement(18801. .18936)
                     /rpt_family="FLAM_C"
                     /evidence=not_experimental
     repeat_region   complement(19119. .19294)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   19346. .19541
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   19687. .19844
                     /rpt_family="LlMEc"
                     /evidence=not_experimental
     repeat_region   complement(19954. .20081)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   20082. .20106
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   20423. .20566
                     /rpt_family="FRAM"
                     /evidence=not_experimental
     repeat_region   20567. .20636
                     /rpt_family="GA-rich"
                     /evidence=not_experimental
     repeat_region   complement(20973. .21282)
                     /rpt_family="AluY"
                     /evidence=not_experimental
     repeat_region   22379. .22724
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   22933. .23019
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   23044. .23270
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   complement(23275. .23586)
                     /rpt_family="AluSgl"
                     /evidence=not_experimental
     repeat_region   complement(23946. .24035)
                     /rpt_family="MER94"
                     /evidence=not_experimental
```

```
     repeat_region   24491. .24777
                     /rpt_family="AluY"
                     /evidence=not_experimental
     repeat_region   24782. .24794
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   24795. .24909
                     /rpt_family="L1PA4"
                     /evidence=not_experimental
     repeat_region   24910. .24939
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(24981. .25276)
                     /rpt_family="AluJb"
                     /evidence=not_experimental
     repeat_region   complement(27879. .28173)
                     /rpt_family="AluSp"
                     /evidence=not_experimental
     repeat_region   complement(28598. .29054)
                     /rpt_family="MER63B"
                     /evidence=not_experimental
     repeat_region   30710. .30738
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(30739. .30993)
                     /rpt_family="AluJb"
                     /evidence=not_experimental
     repeat_region   30994. .31032
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(31408. .31724)
                     /rpt_family="AluJo"
                     /evidence=not_experimental
     repeat_region   complement(31949. .32254)
                     /rpt_family="AluSg"
                     /evidence=not_experimental
     repeat_region   32336. .32362
                     /rpt_family="AT_rich"
                     /evidence=not_experimental
     repeat_region   complement(32463. .32565)
                     /rpt_family="L2"
                     /evidence=not_experimental
     repeat_region   32685. .32898
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   32995. .33020
                     /rpt_family="(TAAA)n"
                     /evidence=not_experimental

  Query Match             28.2%;  Score 487.6;  DB 9;  Length 159712;
  Best Local Similarity   48.0%;  Pred. No. 5.5e-108;
  Matches  484;  Conservative  210;  Mismatches  293;  Indels   22;  Gaps    3;

Qy     744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
           :||:||: |: : ||:||:||::: ||:||:||: | || ||||| :  :| ||:||  |
Db 138232 AGAAGAAGTTCCTAAGCAGCAAAGCATTCAAGAGATAACTTGGGTGTCATTAAAGGCACT 138173

Qy     804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
           ::: ||:||::| ||:|| ||:||:||::: :  ||:||::  || || ||: |:|  ||
Db 138172 CAGTTTTATAAGGGAAGCAGAGCATAAAACTTTGGAAAATTCGCAGCCTGACTATGCAAT 138113

Qy     864 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
           : |:||:||:||:|| ||::: | :|  ||:||:||::  || || ||:||:||:| : |
Db 138112 AGAAAAGAAAAACCCATTTTCTGGGGAGAAATTCAAGCCGGCTGCAGAAATTTGCACAAG 138053

Qy     924 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 983
           :| : |:||::  ||:|| ||:|| ||:||: |:|| ||:||: |:::  || | ||:||
Db 138052 TAGCAAGGAGCCAAATGTTAATCCCCAGGACCATGGGGAAAATGTCTCCAGGCCATGTCA 137993

Qy     984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnc---cnmgnmgnaartggtt 1040
           ::|  : :  ||:||: :  |:  ::: :  || |||:| |   | :| :  ||:|||||
Db 137992 GAGACCTTTACAGCAGCCCCTCCCATCATAAGCCTGGAGGCCTACGAGGAAAAAGTGGTT 137933

Qy    1041 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 1100
           :  :||  | || || || :: :|  |:||:|| ||:|| :| ||::| ||  | ||:||
Db 137932 TCATGGGCCTGGCCCAGGGTCCCTGGGCTGTGTGCAGCCTAGGGACTTGGTGACCTGTGT 137873

Qy    1101 nccngtnaaywsngcngtngcnwsngargg----------------ngcnwsnccnaa 1142
           || |  | : : || || ||      :||                  |   :  |  |
Db 137872 CCCAGCCACTCCAGCCGTGGCTGAAAGGGGCCAGTGTACAGCTTGGGCTGTGGCTTCAGA 137813

Qy    1143 rccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargt 1202
           : | |||||::| ||  : || || ||:|| || || ||  |:||::: :| ||:||:||
Db 137812 GGCTTGGCAGCTTCCACCTGGTGTTGAGCCTGTGGGTGCACAGAAGTCAAGAATTGAGGT 137753

Qy    1203 ntgggarccnccnathmgnttycaraarathtayggnaayccntggatgccnmgncaraa 1262
           |||||:|| |    ::| ||:||:||:|| ||:|| ||: | |||||||| :| ||:||
Db 137752 TTGGGAACCTCTGCCTAGATTTCAGAAGATGTATGGAAATGCCTGGATGCCCAGGCAAAA 137693

Qy    1263 rttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaa 1322
           :||:|| |  || || ||  : :: |||:| || :: || :| |  ||   :||:|  ||
Db 137692 GTTTGCTGCAGGGGTGGGGCCCTCATGGAGAACCTCTGCTAGGGCAGTGAGGAAGGAAAA 137633

Qy    1323 ygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmg 1382
           :|| || |  ||  | || ||::| || || :: || || |   : ::  | || || :|
Db 137632 TGTGGGGTTCGATTCCCCACACAGAGTCCCTACTGGGGCACTGCCTAGTGGAGCTGTGAG 137573

Qy    1383 nmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygt 1442
           :|  : || || :: :  :| :  ||:|| || :| :: || ||::: :| || ||:||
Db 137572 AAGAGGGCCACCATCCTTCAGACCCCAGAATGGTAGATCCACTGACAGCTTGCACCATGT 137513

Qy    1443 nccngaraarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngt 1502
           ||.| :||:::  | ||:|  ||:||:||:||  | ||:|| ||  | | | |: :  |
Db 137512 GCCTGGAAAAACTGCAGACATTCAATGCCAGCCCATGAAAGCAGCCAGGAGGAAGGCTAT 137453

Qy    1503 nccntayaaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytg 1562
           || | :||: |    |  || ||::|  | ||:||  |  | | :||::|| :| ||:
Db 137452 ACCCTGCAAAGCCACAGGGGCAGAGCTGCCCAAGACCATGAGAACCTACCTTTTGCATCA 137393

Qy    1563 ycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttyga 1622
           | :||::| ||:|| :| | :|   | ||: | ||:||:||:|| || :| :  ||:|
Db 137392 GCGTGACCTGGATGTGAGACCTGTAATCAAAGGAGATCATTTTGGAGCTTTAAAATTTGG 137333

Qy    1623 ytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyt 1681
           :||:|| ||  | ||: | || | :||||| || ||  || :| ||:||:||  |:||:|
Db 137332 CTGCCCCACTTGTTTTTGGACTTCCATGGGCCCTGTAACCTCTTTGTTTTGGCTAATTTT 137273

Qy    1682 tyccnttyggnacngcngtnttyacncartgyytntayytncaytgyat 1730
           |:|| ||:|| || |  |  ||: | ||:| ::| ||::  ||:||:||
Db 137272 TCCCATTTGGAACAGTAGCGTTTCCCCAATACCTGTACTCCCATTGTAT 137224


Search completed: July  8, 2002, 18:27:45
Job time: 19721 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 16:20:51 ; Search time 481.64 Seconds
(without alignments)
6170.540 Million cell updates/sec

Title: US-09-997-610-7
Perfect score: 1731
Sequence: 1 atgytnathgargcntaygg.....gyytntayytncaytgyatg 1731

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	471.4	27.2	2590	22	AAH14327	Human cDNA sequenc
c	2	452	26.1	3093	22	AAS03687	Rhesus gene locus:
c	3	452	26.1	9236	22	AAS03689	Rhesus gene locus:
c	4	452	26.1	9236	22	AAS03690	Rhesus gene locus:
c	5	452	26.1	9241	22	AAS03688	Rhesus gene locus:
	6	429.8	24.8	3203	22	AAD05134	Human secreted pro
	7	427.6	24.7	6063	22	AAL03635	Human reproductive
	8	427.2	24.7	22509	22	AAS44505	Human LEKTI DNA cl
	9	427.2	24.7	31529	22	AAS44506	Human LEKTI DNA cl
	10	426.4	24.6	6063	22	AAL03634	Human reproductive
	11	421.4	24.3	1736	22	AAH13678	Human cDNA sequenc
	12	421	24.3	465237	24	ABA90193	Human oestrogen re
	13	409	23.6	1485	23	AAS91984	DNA encoding novel
	14	408.8	23.6	1278	22	AAD05178	Human secreted pro
c	15	400.4	23.1	1580	23	AAS87262	DNA encoding novel
c	16	395.2	22.8	7726	22	ABA21079	Human nervous syst
c	17	393.6	22.7	32127	22	AAI99255	Human excretory re
c	18	393.6	22.7	32127	22	AAI63605	Human kidney relat
c	19	393.6	22.7	72215	22	AAK86832	Human immune/haema
c	20	388.2	22.4	1405	22	AAK70265	Human immune/haema
c	21	386.6	22.3	1405	22	AAK70266	Human immune/haema
	22	386.4	22.3	8387	22	ABA14892	Human nervous syst
	23	386.4	22.3	8387	22	AAK84548	Human immune/haema
	24	384.4	22.2	1621	23	AAS67924	DNA encoding novel
	25	378.8	21.9	3049	22	AAH16637	Human cDNA sequenc
c	26	373.8	21.6	81369	21	AAA97997	Human T gene DNA.
	27	371.6	21.5	1306	23	AAS64678	DNA encoding novel
c	28	361.6	20.9	22756	22	AAS40321	DNA encoding human
c	29	361.6	20.9	22756	22	AAL03921	Human reproductive
	30	355.4	20.5	1946	22	ABA44227	Human breast cell
	31	355.4	20.5	1946	22	ABA54677	Human foetal liver
	32	355.4	20.5	1946	22	ABA24459	Probe #2925 for ge
	33	355.4	20.5	1946	22	AAK02964	Human brain expres
	34	355.4	20.5	1946	22	AAK28410	Human bone marrow
	35	355.4	20.5	1946	22	AAI12975	Probe #2908 for ge
	36	355.4	20.5	1946	22	AAI34334	Probe #3020 used t
	37	355.4	20.5	1946	22	AAI02893	Probe #2884 used t
c	38	347	20.0	26410	22	AAK70623	Human immune/haema
c	39	346.8	20.0	32986	22	AAK69758	Human immune/haema
c	40	346.8	20.0	32986	22	AAK84629	Human immune/haema
	41	346.4	20.0	923	23	AAS66623	DNA encoding novel
	42	341.4	19.7	50000	24	AAD26400	Human glutamate re
	43	341.4	19.7	50000	24	AAD26437	Human GRM3 gene fr
c	44	337.4	19.5	57728	22	AAC87588	Human 9p11 chromos
	45	336.6	19.4	2197	22	AAH18389	Human cDNA sequenc

ALIGNMENTS

RESULT 1
AAH14327
ID AAH14327 standard; cDNA; 2590 BP.
XX
AC AAH14327;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:11697.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.

```
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 11697; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2590 BP; 722 A; 526 C; 670 G; 672 T; 0 other;

  Query Match            27.2%;  Score 471.4;  DB 22;  Length 2590;
  Best Local Similarity  47.0%;  Pred. No. 2e-110;
  Matches  481;  Conservative  213;  Mismatches  293;  Indels   37;  Gaps    3;

Qy     744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
           :||:||:||::: ||:||:||::: ||:||:||:|| |  ||||| :| :| | :|| ||
Db     608 agaagaaatttctaagcagcaaagcattcaagaggtgagttgggtactgttaagagcatt 667

Qy     804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
           ::: ||:| ::| ||: | ||:||:||::  :: ||:||::|  | || ||: |:|  ||
Db     668 cagtttcaaaagggaaacagagcataaaaattcagaaaatttgtagcctgactatgcaat 727

Qy     864 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
           : |:||:||:||:|  ||::: | :|  ||:||:||::  || || ||:||:||:||: |
Db     728 agaaaagaaaaaccagttttctggggagaaattcaagccagctgcagaaatctgcataag 787

Qy     924 yaaygargarytnaaygtnaayccncargayaaygngaraayathwsntggacntgyca 983
           :| : |: |::  ||:|| ||:|| ||:||: |:|| ||:||: |:::  || | ||:||
Db     788 catcaagaagcctaatgttaatccccaggaccatggggaaaatgtctccaggccatgtca 847

Qy     984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naartggtt 1040
           ::|  : :: ||:||: :  |:  ::: :  ||   |:|  | :| :|    ||:|||||
Db     848 gagaccttcacagcagcccctcctgtcacaggcccagagatccaggaggaaaaaatggtt 907

Qy    1041 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 1100
           :| :||  | || || || :: :| ||:||:|| ||:|| :  ||::| || || ||:||
Db     908 ttatgggccaggaccagggtccctgtgctgtgtgcagcctatagacttggtgccctgtgt 967

Qy    1101 nccngtnaayws----------------------------------ngcngtngcnwsnga 1127
            || |    : :                                  || || || :: ||
Db     968 cccagctgctccagccatggctgaaaggggccaatgtagagctcatgctgtggcttcaga 1027

Qy    1128 rggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraa 1187
           :|| |  :: || ||:|| |||||::| ||    || || ||:|| || ||  |  |:||
Db    1028 gggtggaagccccaagccttggcagcttccacatggtgttgagcctgtaggtacacagaa 1087

Qy    1188 rwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaaygccntg 1247
           ::: :| ||:||:|| |||||:|| ||   ::| ||:||:||:|| ||:|| ||: | ||
Db    1088 gtcaagaattgaggtttgggaacctccacctagatttcagaagatgtatggaaatgcctg 1147

Qy    1248 gatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngt 1307
           |||| | :| ||:||:||:|| |  |  |  ||  : :: |||:| || :: || :| |
Db    1148 gatgtccaggcaaaagtttgctgcagtggcagggccctcatggagaacctctgctagggc 1207

Qy    1308 ngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnws 1367
             | ||:||:|| | :|| ||   |||:|  || | ::| || || :  || ||    ::
Db    1208 aatgcagaagggaattgtggggccggagctcccacgtagagtccctaatggggcattatc 1267

Qy    1368 nwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacnga 1427
            ::  | || || :| :|  : || || :: :  :| :  ||: | ||  | :: || ||
Db    1268 tagtggagctgtgagaagagggccaccgtccttcagaccccagcatggtggatccactga 1327

Qy    1428 ywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngc 1487
           ::: :| || ||:|  || | :||: :  | ||:|| ||:||:||: | || ||:|| ||
Db    1328 cagcttgcaccatgcgcctggaaaagccgcagacactcaatgccagtccgtgaaagcagc 1387

Qy    1488 nggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnat 1547
             | | |||: : || || | :||: |    |  || ||::|  | ||: |  | ||
Db    1388 caggagggaggctgtaccctgcaaggccacaggggcagaactgccaaaggccatgggtcc 1447

Qy    1548 htayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttygg 1607
           : |::| :| ||:   | :||::|  |:|| :| | : | || ||: | ||:||:||:||
Db    1448 ccacctttgcatcagcgtgacctgaatgtgagacctagagtcaaaggagatcattttgg 1507

Qy    1608 ngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccng-tnccnytnt 1666
             || :  ||:||:||:||  | || ||: | || | :|||| || |   || :| |
Db    1508 agctttaaaatttgattgtcctgctggattttggacttgcatgggtcctgcagccccttt 1567

Qy    1667 gyttyggncarttyttyccntty ggnacngcngtnttyacncartgyytntayytncayt 1726
           |:||:|| ||:| : |:|| ||:|| |  || || ||:|| ||:| ::| ||::   |:
Db    1568 gttttggccaatgtctcccatttggaatggctgtatttacccaatacctgtacccctatc 1627

Qy    1727 gyat 1730
           |:||
Db    1628 gtat 1631


RESULT   2
AAS03687/c
ID   AAS03687 standard; DNA; 3093 BP.
XX
AC   AAS03687;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Rhesus gene locus: RHD gene deletion in Rh negative haplotypes.
XX
KW   Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW   RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW   haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   primer_bind     32..54
FT                   /*tag=  a
FT                   /note= "Binding site of primer rez7"
FT   primer_bind     complement (3034..3054)
FT                   /*tag=  b
FT                   /note= "Binding site of primer rnb31"
XX
PN   WO200132702-A2.
XX
PD   10-MAY-2001.
XX
PF   31-OCT-2000; 2000WO-EP10745.
XX
PR   02-NOV-1999;   99EP-0121686.
PR   31-MAY-2000; 2000EP-0111696.
XX
PA   (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
```

```
PI   Flegei WA,  Wagner FF;
XX
DR   WPI; 2001-291052/30.
XX
PT   New nucieic acid moiecular structure, usefui for detection of common
PT   RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT   and RHCE genes -
XX
PS   Exampie 10; Fig 5; 135pp; Engiish.
XX
CC   The sequence represents the coding sequence of Rhesus gene locus:
CC   RHD gene deietion in Rh negative haplotypes. The Rhesus genes
CC   iocus comprises the RHD, SMP1 and RHCE (ail undefined) genes and/or the
CC   Rhesus box(es), preferabiy the hybrid Rhesus box, the upstream Rhesus box
CC   and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC   chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC   RHD deietion in the common RHD negative hapiotypes. The sequence has
CC   been used to design primers which are usefui for: (1) the specific
CC   detection of the common RHD positive hapiotypes in D-negative
CC   individuals; (2) blood group typing; (3) determining whether a patient
CC   can be transfused with RhD negative blood and whether biood is suitabie
CC   for transfusion to patients who should not be exposed to antigen C; (4)
CC   assessing the risk of a RhD negative mother of conceiving or carrying an
CC   RhD positive foetus. Anti-D antibodies are usefui for treating pregnant
CC   women who are Rhesus D negative, where the foetus is not homozygous for
CC   the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ   Sequence 3093 BP; 891 A; 754 C; 619 G; 829 T; 0 other;

  Query Match            26.1%;  Score 452;  DB 22;  Length 3093;
  Best Local Similarity  47.4%;  Pred. No. 2.4e-105;
  Matches  485;  Conservative  212;  Mismatches  282;  Indeis   45;  Gaps    5;

Qy     744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
           :| :||:||::: ||:||:||::: ||:||:||:|| || ||||| :| :| ||: | ||
Db    1338 AGGAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGACATT 1279

Qy     804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
           ::: ||:||::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||
Db    1278 CAGTTTTATAAGGGAAGCAGAGCATAAAAGTTCGGAAAATTTGCAGCCTGACAATGTGAT 1219

Qy     864 haaraaraaraa--yccnttywsngarggnaarttyaarytngcngcngarathtgyath 921
           : |:||:||:||  :|| ||::: ||:|  ||:||:||::| || || ||:||:||:||
Db    1218 AGAAAAGAAAAATTCCCATTTTCTGAGGAGAAATTCAAGCTGGCTGCAGAAATTTGCATG 1159

Qy     922 tgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgy 981
            |:||: |:||::  ||:|  ||: | ||:||:||:|| ||:||: |:::  || | ||:
Db    1158 AGTAAC-AGGAGCCAAATGCTAATTCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGT 1100

Qy     982 carmgnwsnwsncarcarwsnathaarwsnytng-cntggmgnccnmgnmgnaartggtt 1040
           ||::| :: :    :||: :  |:  ::: :  | |  |  |    :| :  ||:|||||
Db    1099 CAGAGGTCTTTATGGCAACCCCTCCCATCACAGGTCCAGAGGTATCAGGAAAAAATGGTT 1040

Qy    1041 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 1100
           :||: |  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db    1039 TTGTTGGCCAGGCCCGGGGTCCTCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGCAT 980

Qy    1101 nccngtna----------------------------------aywsngcngtngcnwsng 1126
            || |  |                                      ::: || || || :: |
Db     979 CCCAGCCACTCCCAACCATGACTGACGGGAGGCAAGGTAGAGCTTGGGCTGTAGCTTCGG 920

Qy    1127 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 1186
            : | || :: || ||:|| || ||::| || :: || || ||: | |   | ||  |:|
Db     919 GGAGTGCAAGCCCCAAGCCTTGACAGCTTCCATGTGGTGTTGAGACTGCGAGTGCACAGA 860

Qy    1187 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 1246
           |::: :| ||:| :|| ||| |:||  |   ::| ||  |:| :|| | :|| ||: | |
Db     859 AGTCAAGAATTGGGGTTTGGAAACCTTCGCCTAGATTAAAGAGGATGTGCGGAAATGCCT 800

Qy    1247 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 1306
           ||||||| :| ||:||:||:|| |  || |  ||  : :: |||:|  | :: || :| |
Db     799 GGATGCCCAGTCAGAAGTTTGCTGCAGGAGCAGGGCCCTCATGGAGATCCTCTGCCAGGG 740

Qy    1307 tngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnw 1366
             || ||:||:|| ||:|| || |  ||: | || ||::  || || :: || || ||
Db     739 CAGTGCAGAAGGGAAATGTGGGGTCAGAGACCCCACACACAGTCCCTACTGGGGCACCAC 680

Qy    1367 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1426
           : ::  | || || :| :|  : || ||       :| :  ||:|| || :| :: || |
Db     679 CTAGTGGAGCTGTGAGAAGAGGTCCTCC-------AGACCCCAGAATGGTAGATCCACCG 627

Qy    1427 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1486
           |::: :| || | :|| || | :||: :  | ||:|| ||:||:||:||  | ||:|| |
Db     626 ACAGCTTGCACCGTGTACCTGGAAAAGCTGCAGACACTCAATGCCAGCCCATGAAAGCAG 567

Qy    1487 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1546
           | |  | |||: : || || | :||:       |  || ||::|  |.||:||  | || |
Db     566 CTGAGAGGGAGGCTGTACCCTGCAAAGGTACAGGGGCAGAGCTGCCCAAGACCATGGGAA 507

Qy    1547 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1606
           : |::  :| ||:   | :||::| ||:|| :|   :|| || ||: | ||:||:||:|
Db     506 CCCACCCCTTCCATCTGCGTGACCTGGATGTGAGATGTGGAGTCAAAGGAGATCATTTTG 447

Qy    1607 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytnt 1666
           | || :| :| ||:||:||:|| || || || ||: | ||   :||||| |     || :| |
Db     446 GAGCTTTAAGATTTGACTGCCCCACTGGATTTTGGACTCTCATGGGCCTGTAGCCTCTTT 387

Qy    1667 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1726
           |:||:|  ||:||: |:|| ||:|| |  || || ||:|| ||:||::| ||::  ||:|
Db     386 GTTTTGACCAATTTATCCCATTTGGAATGGCTGTATTTACCCAATGCCTGTACCCCCATT 327

Qy    1727 gyat 1730
           |:||
Db     326 GTAT 323


RESULT   3
AAS03689/c
ID   AAS03689 standard; DNA; 9236 BP.
XX
AC   AAS03689;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Rhesus gene locus: upstream Rhesus box of D-positives.
XX
KW   Rhesus box; RHD positive; sequence iength polymorphism; SSP; RHD; SMP1;
KW   RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW   haemoiytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS   Homo sapiens.
XX
PN   WO200132702-A2.
XX
PD   10-MAY-2001.
XX
PF   31-OCT-2000; 2000WO-EP10745.
XX
PR   02-NOV-1999;   99EP-0121686.
PR   31-MAY-2000; 2000EP-0111696.
XX
PA   (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PI   Fiegei WA,  Wagner FF;
XX
DR   WPI; 2001-291052/30.
XX
PT   New nucleic acid molecular structure, useful for detection of common
PT   RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT   and RHCE genes -
XX
PS   Disciosure; Fig 9; 135pp; Engiish.
XX
CC   The sequence represents the coding sequence of Rhesus gene iocus:
CC   upstream Rhesus box of D positives. The Rhesus genes iocus
```

```
CC   comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC   Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC   and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC   chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC   RHD deletion in the common RHD negative haplotypes. The sequence has
CC   been used to design primers which are useful for: (1) the specific
CC   detection of the common RHD positive haplotypes in D-negative
CC   individuals; (2) blood group typing; (3) determining whether a patient
CC   can be transfused with RhD negative blood and whether blood is suitable
CC   for transfusion to patients who should not be exposed to antigen C; (4)
CC   assessing the risk of a RhD negative mother of conceiving or carrying an
CC   RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC   women who are Rhesus D negative, where the foetus is not homozygous for
CC   the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ   Sequence 9236 BP; 2467 A; 2319 C; 2000 G; 2450 T; 0 other;


  Query Match             26.1%;  Score 452;  DB 22;  Length 9236;
  Best Local Similarity   47.4%;  Pred. No. 6.3e-105;
  Matches  485;  Conservative  212;  Mismatches  282;  Indels   45;  Gaps    5;

Qy     744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
           :| :||:||::: ||:||:||::: ||:||:||:|| || ||||| :| :| ||: | ||
Db    6014 AGGAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGACATT 5955

Qy     804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
           ::: ||:||::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||
Db    5954 CAGTTTTATAAGGGAAGCAGAGCATAAAAGTTCGGAAAATTTGCAGCCTGACAATGTGAT 5895

Qy     864 haaraaraaraa--yccnttywsngarggnaarttyaarytngcngcngarathtgyath 921
           : |:||:||:||  :|| ||::: ||:|  ||:||:||::| || || ||:||:||:||
Db    5894 AGAAAAGAAAAATTCCCATTTTCTGAGGAGAAATTCAAGCTGGCTGCAGAAATTTGCATG 5835

Qy     922 tgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgy 981
            |:||: |:||::  ||:|  ||: | ||:||:||:|| ||:||: |:::  || | ||:
Db    5834 AGTAAC-AGGAGCCAAATGCTAATTCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGT 5776

Qy     982 carmgnwsnwsncarcarwsnathaarwsnytng-cntggmgnccnmgnmgnaartggtt 1040
           ||::| :: :    :||: :  |:  ::: :  | |  |  |    :| :  ||:|||||
Db    5775 CAGAGGTCTTTATGGCAACCCCTCCCATCACAGGTCCAGAGGTATCAGGAAAAAATGGTT 5716

Qy    1041 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 1100
           :||: |  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db    5715 TTGTTGGCCAGGCCCGGGGTCCTCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGCAT 5656

Qy    1101 nccngtna----------------------------------aywsngcngtngcnwsng 1126
            || |  |                                   ::: || || || :: |
Db    5655 CCCAGCCACTCCCAACCATGACTGACGGGAGGCAAGGTAGAGCTTGGGCTGTAGCTTCGG 5596

Qy    1127 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 1186
            : | || :: || ||:|| || ||::| || :: || || ||: | |   | ||  |:|
Db    5595 GGAGTGCAAGCCCCAAGCCTTGACAGCTTCCATGTGGTGTTGAGACTGCGAGTGCACAGA 5536

Qy    1187 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 1246
           |::: :| ||:| :|| ||| |:||  |   ::| ||  |:| :|| | :|| ||: | |
Db    5535 AGTCAAGAATTGGGGTTTGGAAACCTTCGCCTAGATTAAAGAGGATGTGCGGAAATGCCT 5476

Qy    1247 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 1306
           ||||||| :| ||:||:||:|| |  || |  ||  : :: |||:|  | :: || :| |
Db    5475 GGATGCCCAGTCAGAAGTTTGCTGCAGGAGCAGGGCCCTCATGGAGATCCTCTGCCAGGG 5416

Qy    1307 tngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnw 1366
             || ||:||:|| ||:|| || |  ||: | || ||::  || || :: || || ||
Db    5415 CAGTGCAGAAGGGAAATGTGGGGTCAGAGACCCCACACACAGTCCCTACTGGGGCACCAC 5356

Qy    1367 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1426
           : ::  | || || :| :|  : || ||       :| :  ||:|| || :| :: || |
Db    5355 CTAGTGGAGCTGTGAGAAGAGGTCCTCC-------AGACCCCAGAATGGTAGATCCACCG 5303

Qy    1427 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1486
           |::: :| || | :|| || | :||: :  | ||:|| ||:||:||:||  | ||:|| |
Db    5302 ACAGCTTGCACCGTGTACCTGGAAAAGCTGCAGACACTCAATGCCAGCCCATGAAAGCAG 5243

Qy    1487 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1546
           | |  | |||: : || || | :||:      |  || ||::|  | ||:||  | || |
Db    5242 CTGAGAGGGAGGCTGTACCCTGCAAAGGTACAGGGGCAGAGCTGCCCAAGACCATGGGAA 5183

Qy    1547 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1606
            : |::  :| ||:   | :||::| ||:|| :|   :|| || ||: | ||:||:||:|
Db    5182 CCCACCCCTTCCATCTGCGTGACCTGGATGTGAGATGTGGAGTCAAAGGAGATCATTTTG 5123

Qy    1607 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytnt 1666
           | || :| :| ||:||:||:|| || || ||: | ||   :|||||  |     || :| |
Db    5122 GAGCTTTAAGATTTGACTGCCCCACTGGATTTTGGACTCTCATGGGCCTGTAGCCTCTTT 5063

Qy    1667 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1726
           |:||:|  ||:||: |:|| ||:|| |  || || ||:|| ||:||::| ||::  ||:|
Db    5062 GTTTTGACCAATTTATCCCATTTGGAATGGCTGTATTTACCCAATGCCTGTACCCCCATT 5003

Qy    1727 gyat 1730
           |:||
Db    5002 GTAT 4999


RESULT   4
AAS03690/c
ID   AAS03690 standard; DNA; 9236 BP.
XX
AC   AAS03690;
XX
DT   29-AUG-2001  (first entry)
XX
DE   Rhesus gene locus: downstream Rhesus box of D-positives.
XX
KW   Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW   RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW   haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS   Homo sapiens.
XX
PN   WO200132702-A2.
XX
PD   10-MAY-2001.
XX
PF   31-OCT-2000; 2000WO-EP10745.
XX
PR   02-NOV-1999;   99EP-0121686.
PR   31-MAY-2000; 2000EP-0111696.
XX
PA   (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PI   Flegel WA,  Wagner FF;
XX
DR   WPI; 2001-291052/30.
XX
PT   New nucleic acid molecular structure, useful for detection of common
PT   RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT   and RHCE genes -
XX
PS   Disclosure; Fig 10; 135pp; English.
XX
CC   The sequence represents the coding sequence of Rhesus gene locus:
CC   downstream Rhesus box of D positives. The Rhesus genes locus
CC   comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC   Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus box
CC   and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC   chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC   RHD deletion in the common RHD negative haplotypes. The sequence has
CC   been used to design primers which are useful for: (1) the specific
CC   detection of the common RHD positive haplotypes in D-negative
CC   individuals; (2) blood group typing; (3) determining whether a patient
CC   can be transfused with RhD negative blood and whether blood is suitable
CC   for transfusion to patients who should not be exposed to antigen C; (4)
CC   assessing the risk of a RhD negative mother of conceiving or carrying an
CC   RhD positive foetus. Anti-D antibodies are useful for treating pregnant
```

CC women who are Rhesus D negative, where the fDetus is not homozygous for
CC the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ Sequence 9236 BP; 2467 A; 2319 C; 2000 G; 2450 T; 0 other;

Query Match 26.1%; Score 452; DB 22; Length 9236;
Best Local Similarity 47.4%; Pred. No. 6.3e-105;
Matches 485; Conservative 212; Mismatches 282; Indels 45; Gaps 5;

Qy 744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
Db 6014 AGGAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGACATT 5955

Qy 804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
Db 5954 CAGTTTTATAAGGGAAGCAGAGCATAAAAGTTCGGAAAATTTGCAGCCTGACAATGTGAT 5895

Qy 864 haaraaraaraa--yccnttywsngarggnaarttyaarytngcngcngarathtgyath 921
Db 5894 AGAAAAGAAAAATTCCCATTTTCTGAGGAGAAATTCAAGCTGGCTGCAGAAATTTGCATG 5835

Qy 922 tgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgy 981
Db 5834 AGTAAC-AGGAGCCAAATGCTAATTCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGT 5776

Qy 982 carmgnwsnwsncarcarwsnathaarwsnytng-cntggmgnccnmgnmgnaartggtt 1040
Db 5775 CAGAGGTCTTTATGGCAACCCCTCCCATCACAGGTCCAGAGGTATCAGGAAAAAATGGTT 5716

Qy 1041 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 1100
Db 5715 TTGTTGGCCAGGCCCGGGGTCCTCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGCAT 5656

Qy 1101 nccngtna----------aywsngcngtngcnwsng 1126
Db 5655 CCCAGCCACTCCCAACCATGACTGACGGGAGGCAAGGTAGAGCTTGGGCTGTAGCTTCGG 5596

Qy 1127 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 1186
Db 5595 GGAGTGCAAGCCCCAAGCCTTGACAGCTTCCATGTGGTGTTGAGACTGCGAGTGCACAGA 5536

Qy 1187 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 1246
Db 5535 AGTCAAGAATTGGGGTTTGGAAACCTTCGCCTAGATTAAAGAGGATGTGCGGAAATGCCT 5476

Qy 1247 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 1306
Db 5475 GGATGCCCAGTCAGAAGTTTGCTGCAGGAGCAGGGCCCTCATGGAGATCCTCTGCCAGGG 5416

Qy 1307 tngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnw 1366
Db 5415 CAGTGCAGAAGGGAAATGTGGGGTCAGAGACCCCACACACAGTCCCTACTGGGGCACCAC 5356

Qy 1367 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1426
Db 5355 CTAGTGGAGCTGTGAGAAGAGGTCCTCC-------AGACCCCAGAATGGTAGATCCACCG 5303

Qy 1427 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1486
Db 5302 ACAGCTTGCACCGTGTACCTGGAAAAGCTGCAGACACTCAATGCCAGCCCATGAAAGCAG 5243

Qy 1487 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1546
Db 5242 CTGAGAGGGAGGCTGTACCCTGCAAAGGTACAGGGGCAGAGCTGCCCAAGACCATGGGAA 5183

Qy 1547 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1606
Db 5182 CCCACCCCTTCCATCTGCGTGACCTGGATGTGAGATGTGGAGTCAAAGGAGATCATTTTG 5123

Qy 1607 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytnt 1666
Db 5122 GAGCTTTAAGATTTGACTGCCCCACTGGATTTTGGACTCTCATGGGCCTGTAGCCTCTTT 5063

Qy 1667 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1726
Db 5062 GTTTTGACCAATTTATCCCATTTGGAATGGCTGTATTTACCCAATGCCTGTACCCCCATT 5003

Qy 1727 gyat 1730
Db 5002 GTAT 4999

RESULT 5
AAS03688/c
ID AAS03688 standard; DNA; 9241 BP.
XX
AC AAS03688;
XX
DT 29-AUG-2001 (first entry)
XX
DE Rhesus gene locus: hybrid Rhesus box of RHD negatives.
XX
KW Rhesus box; RHD positive; sequence length polymorphism; SSP; RHD; SMP1;
KW RHCE; Rh negative; blood group typing; blood transfusion; antigen C;
KW haemolytic disease of the newborn; chromosome 1 p34.1-p36; ds.
XX
OS Homo sapiens.
XX
PN WO200132702-A2.
XX
PD 10-MAY-2001.
XX
PF 31-OCT-2000; 2000WD-EP10745.
XX
PR 02-NOV-1999; 99EP-0121686.
PR 31-MAY-2000; 2000EP-0111696.
XX
PA (DRKB-) DRK BLUTSPENDEDIENST BADEN WUERTTEMBERG.
XX
PI Flegel WA, Wagner FF;
XX
DR WPI; 2001-291052/30.
XX
PT New nucleic acid molecular structure, useful for detection of common
PT RHD positive haplotypes in D-negative individuals, comprises RHD, SMP1
PT and RHCE genes -
XX
PS Disclosure; Fig 8; 135pp; English.
XX
CC The sequence represents the coding sequence of Rhesus gene locus:
CC hybrid Rhesus box of RHD negatives. The Rhesus genes locus
CC comprises the RHD, SMP1 and RHCE (all undefined) genes and/or the
CC Rhesus box(es), preferably the hybrid Rhesus box, the upstream Rhesus bDx
CC and/or the downstream Rhesus box. The RHD and RHCE genes are located at
CC chromosome 1 p34.1-p36. Rhesus box flanks the breakpoint region of the
CC RHD deletion in the common RHD negative haplotypes. The sequence has
CC been used to design primers which are useful for: (1) the specific
CC detection of the common RHD positive haplotypes in D-negative
CC individuals; (2) blood group typing; (3) determining whether a patient
CC can be transfused with RhD negative blood and whether blood is suitable
CC for transfusion to patients who should not be exposed to antigen C; (4)
CC assessing the risk of a RhD negative mother of conceiving or carrying an
CC RhD positive foetus. Anti-D antibodies are useful for treating pregnant
CC women who are Rhesus D negative, where the foetus is not homozygous for
CC the RHD gene to treat or prevent haemolytic disease of the newborn.
XX
SQ Sequence 9241 BP; 2454 A; 2320 C; 2014 G; 2453 T; 0 other;

Query Match 26.1%; Score 452; DB 22; Length 9241;
Best Local Similarity 47.4%; Pred. No. 6.3e-105;
Matches 485; Conservative 212; Mismatches 282; Indels 45; Gaps 5;

Qy 744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803

```
Db   6014 AGGAGAAATTTCTAAGCAGCAAAGCATTCAAGAGGTGACTTGGGTGCTGTTAAAGACATT 5955

Qy    804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
          ::: ||:||::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||
Db   5954 CAGTTTTATAAGGGAAGCAGAGCATAAAAGTTCGGAAAATTTGCAGCCTGACAATGTGAT 5895

Qy    864 haaraaraaraa--yccnttywsngarggnaarttyaarytngcngcngarathtgyath 921
          : |:||:||:||  :|| ||::: ||:|  ||:|||:||::| || || ||:||:||:||
Db   5894 AGAAAAGAAAAATTCCCATTTTCTGAGGAGAAATTCAAGCTGGCTGCAGAAATTTGCATG 5835

Qy    922 tgyaaygargarytnaaygtnaayccncargayaaygggngaraayathwsntggacntgy 981
           |:||: |:||::  ||:|  ||: | ||:||:||:|| ||:||: |:::  || | ||:
Db   5834 AGTAAC-AGGAGCCAAATGCTAATTCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGT 5776

Qy    982 carmgnwsnwsncarcarwsnathaarwsnytng-cntggmgnccnmgnmgnaartggtt 1040
          ||::| :: :    :||: : |:  ::: :  | |  |  |    :| :  ||:|||||
Db   5775 CAGAGGTCTTTATGGCAACCCCTCCCATCACAGGTCCAGAGGTATCAGGAAAAAATGGTT 5716

Qy   1041 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 1100
          :||: |  | || || || :: :  ||:||:|| ||:|| :| ||:::| || || ||: |
Db   5715 TTGTTGGCCAGGCCCGGGGTCCTCATGCTGTGTGCAGCCTAGGGACTTGGTGCCCTGCAT 5656

Qy   1101 nccngtna----------------------------------aywsngcngtngcnwsng 1126
           || |  |                                   ::: || || || :: |
Db   5655 CCCAGCCACTCCCAACCATGACTGACGGGAGGCAAGGTAGAGCTTGGGCTGTAGCTTCGG 5596

Qy   1127 arggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaara 1186
          : | || :: || ||:|| || ||::| || :: || || ||: | |   | ||  |:|
Db   5595 GGAGTGCAAGCCCCAAGCCTTGACAGCTTCCATGTGGTGTTGAGACTGCGAGTGCACAGA 5536

Qy   1187 arwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 1246
          |::: :| ||:| :|| ||| |:||  |   ::| ||  |:| :|| | :|| ||: | |
Db   5535 AGTCAAGAATTGGGGTTTGGAAACCTTCGCCTAGATTAAAGAGGATGTGCGGAAATGCCT 5476

Qy   1247 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 1306
          ||||||| :| ||:||:||:|| |  || |  ||  : :: |||:|  | :: || :| |
Db   5475 GGATGCCCAGTCAGAAGTTTGCTGCAGGAGCAGGGCCCTCATGGAGATCCTCTGCCAGGG 5416

Qy   1307 tngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnw 1366
            || ||:||:|| ||:|| || |  ||: | || ||::  || || :: || || ||
Db   5415 CAGTGCAGAAGGGAAATGTGGGGTCAGAGACCCCACACACAGTCCCTACTGGGGCACCAC 5356

Qy   1367 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1426
          : ::  | || || :| :|  : || ||       :| :  ||:|| || :| :: || |
Db   5355 CTAGTGGAGCTGTGAGAAGAGGTCCTCC-------AGACCCCAGAATGGTAGATCCACCG 5303

Qy   1427 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1486
          |::: :| || | :|| || | :||: :  | ||:|| ||:|||:||:||  | ||:|| |
Db   5302 ACAGCTTGCACCGTGTACCTGGAAAAGCTGCAGACACTCAATGCCAGCCCATGAAAGCAG 5243

Qy   1487 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1546
          | |  | |||: : || || | :||:       |  || ||::|  | ||:||  | || |
Db   5242 CTGAGAGGGAGGCTGTACCCTGCAAAGGTACAGGGGCAGAGCTGCCCAAGACCATGGGAA 5183

Qy   1547 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1606
           : |::  :| ||:   | :||::| ||:|| :|   :|| || ||: | ||:||:||:|
Db   5182 CCCACCCCTTCCATCTGCGTGACCTGGATGTGAGATGTGGAGTCAAAGGAGATCATTTTG 5123

Qy   1607 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggncngtnccnytnt 1666
          | || :| :| ||:||:||:|| || || ||: | ||   :||||| |     || :| |
Db   5122 GAGCTTTAAGATTTGACTGCCCCACTGGATTTTGGACTCTCATGGGCCTGTAGCCTCTTT 5063

Qy   1667 gyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncayt 1726
          |:||:|  ||:||: |:|| ||:|| |  || || ||:|| ||:||::| ||::  ||:|
Db   5062 GTTTTGACCAATTTATCCCATTTGGAATGGCTGTATTTACCCAATGCCTGTACCCCCATT 5003

Qy   1727 gyat 1730
          |:||
Db   5002 GTAT 4999


RESULT   6
AAD05134
```

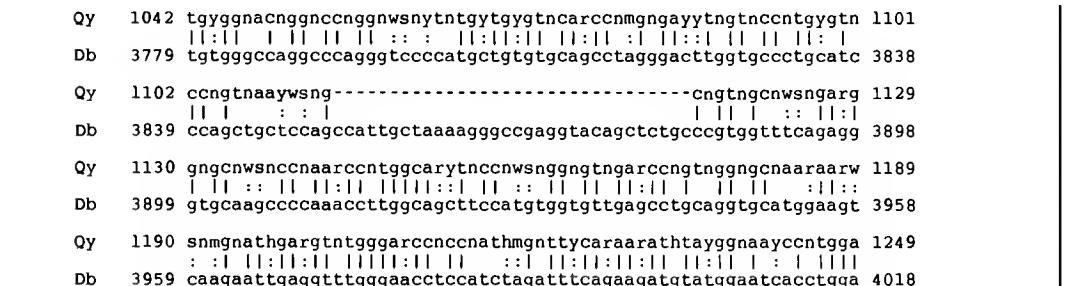
```
ID   AAD05134 standard; cDNA; 3203 BP.
XX
AC   AAD05134;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 14 cDNA clone HISBF60, SEQ ID NO:24.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             153..356
FT                   /*tag= a
FT                   /product= "Human secreted protein"
FT   sig_peptide     153..215
FT                   /*tag= b
FT   mat_peptide     216..353
FT                   /*tag= c
FT                   /product= "Human mature secreted protein"
XX
PN   WO200134769-A2.
XX
PD   17-MAY-2001.
XX
PF   01-NOV-2000; 2000WO-US30040.
XX
PR   05-NOV-1999;   99US-0163580.
PR   30-JUN-2000; 2000US-0215130.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Wei P,  Baker KP,  Fiscella M;
XX
DR   WPI; 2001-308781/32.
DR   P-PSDB; AAE01245.
XX
PT   New isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition  -
XX
PS   Claim 1; Page 408-409; 519pp; English.
XX
CC   AAD05121-AAD05203 represent cDNAs corresponding to 24 human secreted
CC   protein genes, and AAE01232-AAE01311 represent the proteins they encode.
CC   AAE01312-AAE01340 represent human secreted protein variants or fragments.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   24 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
```

```
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 3203 BP; 868 A; 650 C; 740 G; 943 T; 2 other;


  Query Match             24.8%;  Score 429.8;  DB 22;  Length 3203;
  Best Local Similarity   46.5%;  Pred. No. 1.3e-99;
  Matches  476;  Conservative  211;  Mismatches  297;  Indels   40;  Gaps    6;

Qy     745 gargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntty 804
           ||:||:||::: ||:||:||::: ||:||:||:|| |  ||||| :| :| ||:|| ||:
Db    1639 gaagaaatttctaagcagcaaagcattcaagaggtgatttgggtgctgttaaaagcattt 1698

Qy     805 wsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnath 864
              ||:| ::| ||: |  |:||:||::  :  ||:||::| || || | : |:|    :
Db    1699 cattttaaaagggaaac--agcataaaaatttggaaaatttgcagccagttgatgcagca 1756

Qy     865 aaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtgy 924
            |: |:| :||:|  ||::  ||:|  ||:||:||::| || || ||:||:||:||: |:
Db    1757 gaagagagaaacctgttttttgaggagaaattcaagctggctgcagaaatttgcataagt 1816

Qy     925 aaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycar 984
            |:  :||::| ||:|| ||:|  ||:||:||:|| ||:||: |:::  || | ||:||
Db    1817 tacagggagctgaatgttaatctccaagacaatggggaaaatgtctccagggcatgtcac 1876

Qy     985 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tggtty 1041
           :| :: ::   :||: :  |:  ::: :   |   |:   | :| :| ||:   |||||:
Db    1877 aggtcttcatggcagcccctcccatcacagaccctgaagactaggaggaaaaaatggttt 1936

Qy    1042 tgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtn 1101
           ||:||  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db    1937 tgtgggccaggcccagggtccccatgctgtgtgcagcctagggacttggtgccctgcatc 1996

Qy    1102 ccngtnaaywsngcngtngc------------------------------nwsngarg 1129
           || |    : : ||  | ||                              :: ||:|
Db    1997 ccagctgctccagccattgctaaaagggccgaggtacagctctgcccatggtttcagagg 2056

Qy    1130 gngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarw 1189
           | || :: || ||:|| |||||::| || :: || || ||:|| |  || ||   :||::
Db    2057 gtgcaagccccaaaccttggcagcttccatgtggtgttgagcctgcaggtgcatggaagt 2116

Qy    1190 snmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaaygccntgga 1249
           : :| ||:||:|| |||||:|| ||   ::| ||:||:||:|| ||:|| | : | ||||
Db    2117 caagaattgaggtttgggaacctccatctagatttcagaagatgtatggaatcacctgga 2176

Qy    1250 tgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtng 1309
           |||  :| ||:||:||:||    || |  |   : :: |||:| || :: || :| |  |
Db    2177 tgctcaggcaaaagtttgctccaggggcagagccttcatggagaacctctgctagggcag 2236

Qy    1310 tncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnw 1369
           |   :||:|| ||:|| || | |||:|| || ||::|  | || :: || || ||  : :
Db    2237 tg-tgaagggaaatgtggggttggagcccccacacagaatccctactggggcaccaccta 2295

Qy    1370 snmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayw 1429
           :  | || || :| :| :: || || :: :  :| :  ||:|| || :| :: || ||::
Db    2296 gtggagctgtgcgaagaaggccaccgtccttcagaccccagaatggtagatccaccgaca 2355

Qy    1430 snytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng-cn 1488
           : :|  | | :|  || | :||: : || ||: |  |: |:||:|| || ||:|| | |
Db    2356 gcttgtaccgtgcacctggaaaagccacagaccctaaacgccagcccgtgaaagcagcca 2415

Qy    1489 ggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1548
           ||  ||| :::  | || | : |: |    |  || ||::|  | ||:||    || | :
Db    2416 ggagtggggactataccctgtgaagccacaggggcagagctgcccaagactaagggaacc 2475

Qy    1549 tayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggn 1608
           ||::| :| ||:  :  :||::| ||:|| :| ||:|| || | : | ||:||:||:||
Db    2476 tacctcttgcatcattgtgacctggatgtgagacatggagtcagaggagatcattttgga 2535

Qy    1609 gcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntg 1667
             |  |    ||:||:||:|   | || ||: | || | :||||| |  ||  |  :| ||
Db    2536 acgntataatttgactgcctcgctggattttggacttgcatggggcttgtaacgcctttg 2595

Qy    1668 yttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1727
           :| :|| ||:||: |:|| ||:|| |  || || ||:|| || | ::| ||::  ||:||
Db    2596 ttgtgggcagtttctcccatttggaatggctgtatttacccattacctgtatccccattg 2655

Qy    1728 yatg 1731
           :|||
Db    2656 tatg 2659


RESULT   7
AAL03635
ID   AAL03635 standard; DNA; 6063 BP.
XX
AC   AAL03635;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6323.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
```

PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 6323; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 6063 BP; 1655 A; 1162 C; 1259 G; 1987 T; 0 other;

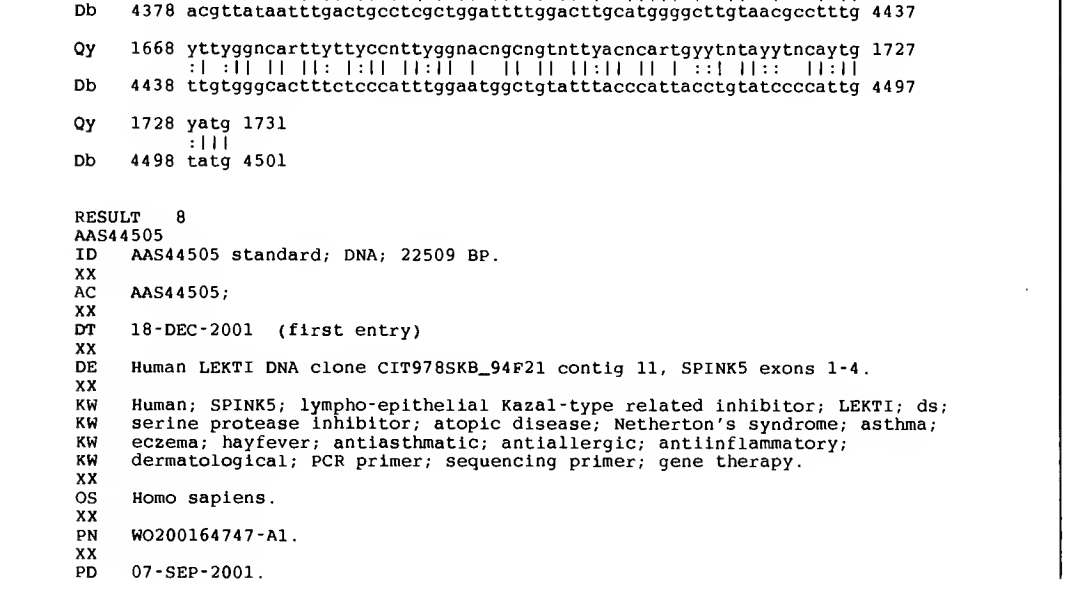
Query Match 24.7%; Score 427.6; DB 22; Length 6063;
Best Local Similarity 46.4%; Pred. No. 8.4e-99;
Matches 475; Conservative 211; Mismatches 298; Indels 40; Gaps 6;

```
Qy    745 gargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntty 804
          ||:||:||::: ||:|||:||::: ||:||:|||:|| |  ||||| :| :| ||:|| ||:
Db   3481 gaagaaatttctaagcagcaaagcattcaagaggtgatttgggtgctgttaaaagcattt 3540

Qy    805 wsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnath 864
            ||:| ::| ||: |  |:|||:||::  :  ||:||::| || || | : |:|    :
Db   3541 cattttaaaagggaaac--agcataaaaatttggaaaatttgcagccagttgatgcagca 3598

Qy    865 aaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtgy 924
            |: |:| :||:|  ||::  ||:|  ||:||:||::| || || ||:||:||:||: |:
Db   3599 gaagagagaaacctgttttttgaggagaaattcaagctggctgcagaaatttgcataagt 3658

Qy    925 aaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycar 984
            |:  :||::| ||:|| ||:|  ||:||:||:|| ||:||: |:::  || | ||:||
Db   3659 tacagggagctgaatgttaatctccaagacaatggggaaaatgtctccagggcatgtcac 3718

Qy    985 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tggtty 1041
          :| :: ::   :||: :  |:  ::: :   |   |:   | :| :| ||:   |||||:
Db   3719 aggtcttcatggcagcccctcccatcacagaccctgaagactaggaggaaaaaatggttt 3778
```



RESULT 8
AAS44505
ID AAS44505 standard; DNA; 22509 BP.
XX
AC AAS44505;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human LEKTI DNA clone CIT978SKB_94F21 contig 11, SPINK5 exons 1-4.
XX
KW Human; SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI; ds;
KW serine protease inhibitor; atopic disease; Netherton's syndrome; asthma;
KW eczema; hayfever; antiasthmatic; antiallergic; antiinflammatory;
KW dermatological; PCR primer; sequencing primer; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200164747-A1.
XX
PD 07-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-GB00897.
XX
PR 02-MAR-2000; 2000GB-0005098.
PR 03-MAR-2000; 2000GB-0005229.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Hovnanian A, Chavanas S, Cookson W, Moffat M, Walley A;
XX
DR WPI; 2001-582149/65.
XX
PT Determining susceptibility to atopic disease or carrier status of
PT Netherton's syndrome in humans by identifying variants of or mutations
PT in SPINK5, a gene encoding lympho-epithelial Kazal-type related
PT inhibitor -
XX
PS Disclosure; Page 88-94; 123pp; English.
XX
CC Sequences AAS44359-AAS44514 represent the SPlNK5 gene, contigs and
CC fragments of a SPlNK5 clone, sequencing primers and PCR primers for
CC SPINK5. SPINK5 encodes lympho-epithelial Kazal-type related inhibitor
CC (LEKTl), a serine protease inhibitor. Susceptibility or predisposition to
CC an atopic disease in a human subject can be detected by screening the
CC genome for one or more polymorphic variants of SPINK5 gene and/or
CC expression of a variant LEKTI protein in a tissue. Carrier status of a
CC subject or development of Netherton's syndrome is diagnosed by screening
CC for the presence of loss-of-function mutations in the SPlNK5 gene. An
CC expression vector comprising a nucleic acid encoding a serine protease
CC inhibitor or its functional fragment can be used in screening for
CC compounds with potential pharmacological activity by determining the
CC serine protease activity of a protein previously identified as a ligand
CC of the LEKTI protein. The atopic diseases include Netherton's Syndrome,
CC asthma, eczema and hayfever.
XX
SQ Sequence 22509 BP; 6725 A; 4291 C; 4284 G; 7209 T; 0 other;

Query Match 24.7%; Score 427.2; DB 22; Length 22509;
Best Local Similarity 47.1%; Pred. No. 3.4e-98;
Matches 472; Conservative 207; Mismatches 303; Indels 21; Gaps 7;



```
Qy   1209 rccnccnathmgnttycaraarathtayggnaayccntggatgccnmgncaraarttygc 1268
          :|| |    ::| | :|   |:||:||:|| ||: | |||||| | :| ||:||:||: |
Db  14666 acctcagcctagatctc---agatatatggaaatgcctggatgtccaggcagaagttttc 14722

Qy   1269 ngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtngg 1328
           || || ||  | :: :: |||:| || :: || :| |  || ||:||: | ||:|| |
Db  14723 tgtaggggtgaggtcctcatggagaacctctgctagggcagtacagaagagaaatgtgga 14782

Qy   1329 ntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnws 1388
           | |||:|| || ||::| || || :: || || ||     :  | || || :| :|
Db  14783 gtaggagcccccacacagagtccctactgg-gcaccacttggtggagctgttagaagaga 14841

Qy   1389 nccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccnga 1448
           ||  | :: :: :|    ||:||  | :| :: || | ::: :| || ||: | || ||
Db  14842 gccttcatcctccagaacccagaatagcagatccacaggtagcttgcaccatatgcctga 14901

Qy   1449 raarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccnta 1508
          :||: : ||  |:|| ||:||:||:|| || ||:|| || || | |||: : || || |
Db  14902 gaaagccacaaacactcagtgccagcctgtgaaggcagctgggagggaggcggtaccctg 14961

Qy   1509 yaaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycayga 1568
          :||: |    |  |  ||::|  | ||:||  | || | : |:::| :| ||:   ||:||
Db  14962 caaagccacagaggtggagttgcccaagaccatgggaacccacctcttgcatcagcatga 15021

Qy   1569 yytngaygtnmgncayggngtnaarmgngaycaytttyggngcnytnmgnttygaytgycc 1628
          ::| ||: | :| ||: |  | ||::   |:||:||:||  | :| :| ||:||:||:|
Db  15022 cctggatatgagacatagattcaaaatatatcattttggaactttaagatttgactgcct 15081

Qy   1629 nacnggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnt 1687
            | |  ||:  || | :|||  | ||  || :| ||:||:|| ||:| : |:|| |
Db  15082 tcctgaattttcgacttgcatggcatctgtagccccttttgttttggccaatatctcccat 15141

Qy   1688 tyggnacngcngtnttyacncartgyytntayytncaytgyat 1730
          |:|| || || || || || ||:|| :| ||::  ||:||:||
Db  15142 ttggaacagctgtattgacccaatggctgtaccccccattgtat 15184
```

RESULT 9
AAS44506
ID AAS44506 standard; DNA; 31529 BP.
XX
AC AAS44506;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human LEKTI DNA clone CIT978SKB_94F21 contig 8, SPINK5 exon 5.
XX
KW Human; SPINK5; lympho-epithelial Kazal-type related inhibitor; LEKTI; ds;
KW serine protease inhibitor; atopic disease; Netherton's syndrome; asthma;
KW eczema; hayfever; antiasthmatic; antiallergic; antiinflammatory;
KW dermatological; PCR primer; sequencing primer; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200164747-A1.
XX
PD 07-SEP-2001.
XX
PF 02-MAR-2001; 2001WO-GB00897.
XX
PR 02-MAR-2000; 2000GB-0005098.
PR 03-MAR-2000; 2000GB-0005229.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Hovnanian A, Chavanas S, Cookson W, Moffat M, Walley A;
XX
DR WPI; 2001-582149/65.
XX
PT Determining susceptibility to atopic disease or carrier status of
PT Netherton's syndrome in humans by identifying variants of or mutations
PT in SPINK5, a gene encoding lympho-epithelial Kazal-type related
PT inhibitor -
XX
PS Disclosure; Page 95-103; 123pp; English.
XX
CC Sequences AAS44359-AAS44514 represent the SPINK5 gene, contigs and
CC fragments of a SPINK5 clone, sequencing primers and PCR primers for
CC SPINK5. SPINK5 encodes lympho-epithelial Kazal-type related inhibitor
CC (LEKTI), a serine protease inhibitor. Susceptibility or predisposition to
CC an atopic disease in a human subject can be detected by screening the
CC genome for one or more polymorphic variants of SPINK5 gene and/or
CC expression of a variant LEKTI protein in a tissue. Carrier status of a
CC subject or development of Netherton's syndrome is diagnosed by screening
CC for the presence of loss-of-function mutations in the SPINK5 gene. An
CC expression vector comprising a nucleic acid encoding a serine protease
CC inhibitor or its functional fragment can be used in screening for
CC compounds with potential pharmacological activity by determining the
CC serine protease activity of a protein previously identified as a ligand
CC of the LEKTI protein. The atopic diseases include Netherton's Syndrome,
CC asthma, eczema and hayfever.
XX
SQ Sequence 31529 BP; 9686 A; 5798 C; 5934 G; 10111 T; 0 other;

Query Match 24.7%; Score 427.2; DB 22; Length 31529;
Best Local Similarity 47.1%; Pred. No. 4.6e-98;
Matches 472; Conservative 207; Mismatches 303; Indels 21; Gaps 7;

```
Qy    744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
          :|||:||:||::: ||:||:||::: ||:||:||: | || |||||  :| :| ||:|| ||
Db  14187 agaagaaatttctaagcagcagagcattcaagagataacttgggtgctcttaaaggcatt 14246

Qy    804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
          ::: ||: |::| ||:|| ||:||:||::: :  ||:||::| || || ||:| :|| ||
Db  14247 cagctttttaagggaagcagaacataaaagtttggaaaatttgcagcctgacagtgtgat 14306

Qy    864 haaraara--araayccnttywsngarggnaarttyaarytngcngcngarathtgyath 921
          :|||:||:|  |:||:|| ||::: ||:|  ||:||:||::  |  ||  |:||:||:||:
Db  14307 aaaaaaaaagaaaatccatttttctgaggagaaattcaagccagttgcataaatttgcata 14366

Qy    922 tgyaaygargarytnaaygtnaaycccncargayaaygggngaraayathwsntggacntgy 981
          |:| : |:||::   |:|| ||:|| ||:||:||:|| ||:||: |:::   ||  | :
Db  14367 agtagcaaggagcctgatgttaatccccaagacaatgg-gaaaatgtctccaagagatat 14425

Qy    982 carmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tgg 1038
          ||::|  : :    :||: :  |:  ::  :  ||   |:|    :| :| ||:   |||
Db  14426 cagagacctttgtggcagcccctcccattacatgcccagaggtttaggaggaaaaaatgg 14485

Qy   1039 ttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgy 1098
          ||:| : |  | ||    || :: :| ||:||:|| ||:||  | | :   |   |
Db  14486 ttctctaggccaggcttagggtccctctgctgtgtgcagccctgcgttgaagacactcca 14545

Qy   1099 gtnccng----------tnaaywsngcngtngcnwsngarggngcnwsnccnaarccntg 1148
          ||  | |          |   ::: ||  | || :: ||:|| || :: |  ||:|| ||
Db  14546 gtggctgaaggggccatggcttgggctatggcttccgagggtgcaagcctgaagccttg 14605

Qy   1149 gcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntggga 1208
           ||:|| ||    || || ||:|| ||  | ||  |:||::: :| | :||:|| ||  |
Db  14606 acagcttccacatggtgttgagcctgtgagtgcacagaagtcaagaactgaggtttgaaa 14665

Qy   1209 rccnccnathmgnttycaraarathtayggnaayccntggatgccnmgncaraarttygc 1268
          :|| |    ::| | :|   |:||:||:|| ||: | |||||| | :| ||:||:||: |
Db  14666 acctcagcctagatctc---agatatatggaaatgcctggatgtccaggcagaagttttc 14722

Qy   1269 ngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtngg 1328
           || || ||  | :: :: |||:| || :: || :| |  || ||:||: | ||:|| |
Db  14723 tgtaggggtgaggtcctcatggagaacctctgctagggcagtacagaagagaaatgtgga 14782

Qy   1329 ntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnws 1388
           | |||:|| || ||::| || || :: || || ||     :  | || || :| :|
Db  14783 gtaggagcccccacacagagtccctactgg-gcaccacttggtggagctgttagaagaga 14841

Qy   1389 nccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccnga 1448
           ||  | :: :: :|    ||:||  | :| :: || | ::: :| || ||: | || ||
```

```
Db  14842 gccttcatcctccagaacccagaatagcagatccacaggtagcttgcaccatatgcctga 14901

Qy   1449 raarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccnta 1508
          :||: : ||  |:|| ||:||:||:|| || ||:|| || || | |||: : || || |
Db  14902 gaaagccacaaacactcagtgccagcctgtgaaggcagctgggagggaggcggtaccctg 14961

Qy   1509 yaaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycayga 1568
          :||: |    |  |  ||::|  | ||:||  | || | : |::| :| ||:   ||:||
Db  14962 caaagccacagaggtggagttgcccaagaccatgggaacccacctcttgcatcagcatga 15021

Qy   1569 yytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgycc 1628
          ::| ||: | :| ||: |  | ||::   |:||:||:||  | :| :| ||:||:||:|
Db  15022 cctggatatgagacatagattcaaaatatatcattttggaactttaagatttgactgcct 15081

Qy   1629 nacnggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnt 1687
            | |  ||:   || | :|||   | ||  || :| ||:||:|| ||:| : |:|| |
Db  15082 tcctgaattttcgacttgcatggcatctgtagccccttttgttttggccaatatctcccat 15141

Qy   1688 tyggnacngcngtnttyacncartgyytntayytncaytgyat 1730
          |:|| || || || || || ||:|| :| ||::  ||:||:||
Db  15142 ttggaacagctgtattgacccaatggctgtacccccattgtat 15184


RESULT  10
AAL03634
ID   AAL03634 standard; DNA; 6063 BP.
XX
AC   AAL03634;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6322.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
```

```
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen
PT   is used in preventing, treating or ameliorating a medical condition  -
XX
PS   Disclosure; SEQ ID NO 6322; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention.
XX
SQ   Sequence 6063 BP; 1656 A; 1163 C; 1262 G; 1982 T; 0 other;

  Query Match             24.6%;  Score 426.4;  DB 22;  Length 6063;
  Best Local Similarity   46.4%;  Pred. No. 1.7e-98;
  Matches  475;  Conservative  210;  Mismatches  299;  Indels   40;  Gaps    6;

Qy     745 gargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntty 804
           ||:||:||::: ||:||:||::: ||:||:||:|| |  ||||| :| :| ||:|| ||:
Db    3481 gaagaaatttctaagcagcaaagcattcaagaggtgatttgggtgctgttaaaagcattt 3540

Qy     805 wsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnath 864
             ||:| ::| ||: |  |:||:||::  :  ||:||::| || || | : |:|    :
Db    3541 cattttaaaagggaaac--agcataaaaatttggaaaatttgcagccagttgatgcagca 3598

Qy     865 aaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtgy 924
            |: |:| :||:|  ||::  ||:|  ||:||:||::| || || ||:||:||:||: |:
Db    3599 gaagagagaaacctgttttttgaggagaaattcaagctggctgcagaaatttgcataagt 3658

Qy     925 aaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycar 984
            |:  :||::| ||:|| ||:|  ||:||:||:|| ||:||: |:::  || | ||:||

Db    3659 tacagggagctgaatgttaatctccaagacaatggggaaaatgtctccagggcatgtcac 3718

Qy     985 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tggtty 1041
           :| :: ::   :||: :  |:  ::: :   |   |:   | :| :| ||:   |||||:
Db    3719 aggtcttcatggcagcccctcccatcacagaccctgaagactaggaggaaaaaatggttt 3778

Qy    1042 tgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtn 1101
           ||:||  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db    3779 tgtgggccaggcccagggtccccatgctgtgtgcagcctagggacttggtgccctgcatc 3838

Qy    1102 ccngtnaaywsng----------------------------------cngtngcnwsngarg 1129
           || |    : : |                                  | || |  :: ||:|
Db    3839 ccagctgctccagccattgctaaaagggccgaggtacagctctgcccgtggtttcagagg 3898

Qy    1130 gngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarw 1189
           | || :: || ||:|| |||||::| || :: || || ||:|| |  || ||   :||::
Db    3899 gtgcaagccccaaaccttggcagcttccatgtggtgttgagcctgcaggtgcatggaagt 3958

Qy    1190 snmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntgga 1249
           : :| ||:||:|| |||||:|| ||   ::| ||:||:||:|| ||:|| | : | ||||
Db    3959 caagaattgaggtttgggaacctccatctagatttcagaagatgtatggaatcacctgga 4018

Qy    1250 tgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtng 1309
           |||  :| ||:||:|||:||    || |  |   : :: |||:| || :: || :| |  |
Db    4019 tgctcaggcaaaagtttgctccaggggcagagccttcatggagaacctctgctagggcag 4078

Qy    1310 tncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnw 1369
           |   :||:|| ||:|| || | |||:|| || ||::|  | || :: || || ||  : :
Db    4079 tg-tgaagggaaatgtggggttggagcccccacacagaatccctactggggcaccaccta 4137

Qy    1370 snmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayw 1429
           :  | || || :  :| :: || || :: :  :| :  ||:|| || :| :: || ||::
Db    4138 gtggagctgtgcaaagaaggccaccgtccttcagaccccagaatggtagatccaccgaca 4197

Qy    1430 snytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng-cn 1488
           : :|  | | :|  || | :||: : || ||: |  |: |:||:|| || ||:|| | |
Db    4198 gcttgtaccgtgcacctggaaaagccacagaccctaaacgccagcccgtgaaagcagcca 4257

Qy    1489 ggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1548
           ||  ||| :::  | || | : |: |    |  || ||::|  | ||:||    || | :
Db    4258 ggagtggggactataccctgtgaagccacaggggcagagctgcccaagactaagggaacc 4317

Qy    1549 tayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggn 1608
           ||::| :| ||:  :  :||::| ||:|| :| ||:|| || | : | ||:||:||:||
Db    4318 tacctcttgcatcattgtgacctggatgtgagacatggagtcagaggagatcattttgga 4377

Qy    1609 gcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntg 1667
            | :|    || ||:||:|   | || ||: | || | :|||||  |  ||  |  :| ||
Db    4378 acgttataattggactgcctcgctggattttggacttgcatggggcttgtaacgcctttg 4437

Qy    1668 yttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytg 1727
           :| :|| || ||: |:|| ||:|| |  || || ||:|| || | ::| ||::  ||:||
Db    4438 ttgtgggcactttctcccatttggaatggctgtatttacccattacctgtatccccattg 4497

Qy    1728 yatg 1731
           :|||
Db    4498 tatg 4501


RESULT  11
AAH13678
ID   AAH13678 standard; cDNA; 1736 BP.
XX
AC   AAH13678;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:10539.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
```

```
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0116126.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   11-JAN-2000; 2000JP-0118776.
PR   02-MAY-2000; 2000JP-0183767.
PR   09-JUN-2000; 2000JP-0241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI   Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 10539; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 1736 BP; 449 A; 372 C; 453 G; 462 T; 0 other;


  Query Match            24.3%;  Score 421.4;  DB 22;  Length 1736;
  Best Local Similarity  46.2%;  Pred. No. 1.1e-97;
  Matches  474;  Conservative  211;  Mismatches  293;  Indels   47;  Gaps    6;

Qy    744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
          :||:||:||::: ||:||:||::: ||:||:||:|| |  ||| | :| :| ||:|| ||
Db    110 agaagaaatttctaagcagcaaagcattcaagaggtgatttggatactgttaaaggcatt 169

Qy    804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
          ::: ||:||::| ||:|| ||:||:||::: :  ||:||::| || |   |: |:|  ||
Db    170 tagttttataagggaagcagagcataaaagtttggaaaatttgcagcattactatgcgat 229

Qy    864 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
          : | ||:||:||:|| ||::: | :|  | :||:||::  || || ||:||:||: |: |
Db    230 agacaagaaaaacccatttctggggagagattcaagccagctgcggaaatttgtgtaag 289

Qy    924 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 983
          :| : |:||::  ||:|| | :|| ||:||: |:|| ||: |: |:::  |  | ||:||
Db    290 tagcaaggagcctaatgttagtccccaagaccatggggaagatgtctccagaccatgtca 349

Qy    984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 1043
          ::|  : :: | :||: :  |:  ::: :  ||   |:     :| :  ||:|||||:||
Db    350 gagaccttcaccacagcccctcctatcacaggcccaga-----aggaaaaagtggttttg 404

Qy   1044 yggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtncc 1103
          :||  | || || || :: :  ||:||:|| ||:|| :| ||: | |  || ||:|| ||
Db    405 tgggcctggtccagggtccccttgctgtgtgcaacctagggatgtggcaccctgtgtccc 464

Qy   1104 ngtnaayws----------------------------------ngcngtngcnwsnga 1127
            |   : :                                   |  || || :: ||
Db    465 agctgctgcttcagctgtggctgaaaggggccaatgtacagctcaggttgtggcctcaga 524

Qy   1128 rggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraa 1187
          :|| |  :: || ||:|  ||||||::  ||    || || ||:|| |  || ||  |:||
Db    525 gggtggaagccccaagctttggcagcgtccacatggtgttgagcctgcaggtgcacagaa 584

Qy   1188 rws-nmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccnt 1246
          :::  :| ||:||:|| |||||:|| |    ::| ||:|   |:|| ||:|| ||:   |
Db    585 gtcaaagaattgaggtttgggaacctctacctagatttc---agatgtatggaaatgtct 641

Qy   1247 ggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgng 1306
          |||||| :| ||:||:||:||  | || || ||  : :: | |:| || :: || :| |
Db    642 agatgcccaggcaaaagtttgctat-ggggtggggccctcatagagaacctctgctaggg 700

Qy   1307 tngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnw 1366
            ||   :||:|| ||:|| || | |||:|| || ||::| || || :: || || |
Db    701 cagtgtggaagggaaatgtggggttggagccccccacacagagtccctactggggcactgc 760

Qy   1367 snwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacng 1426
          : :: :| || || :| :|  : ||    :: :: :| :  ||:|| || :|  : || |
Db    761 ctagtagagctgtgagaagagggccattgtcttctagaccccagaatggcagacccactg 820

Qy   1427 aywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng 1486
          |::: :| || ||:|| || | :||: :  | ||:|| ||: |:||:|| || ||:||
Db    821 acagcttgcaccatgtgcctggaaaagctgcagacactcaacgccagcccgtgaaagcat 880

Qy   1487 cnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1546
          | || | |||: : || || | : |: |    |  || ||::|  | ||:||  | || |
Db    881 ccgggagggaggctgtaccctgccaagccacaggggcagagctacccaagaccatgggaa 940

Qy   1547 thtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1606
          : |::| :| ||:    :||::| ||: | :| | :||  | ||: | ||:|||::|
Db    941 ctcacctcttgcatcagtgtgacctggatttgagacctggaatcaaaggagatcattttg 1000

Qy   1607 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytn 1665
          | || :| :  ||:||:||:|| || |  ||: | || | :|||| || ||  || :
Db   1001 gagctttaacatttgactgccccactgaattctggacttgcatgggacccgtaaccccat 1060

Qy   1666 tgyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncay 1725
          ||:||:|| ||:||: |: | ||:|| |  || || ||:|| ||:| ::|  |:   ||:
Db   1061 tgttttggccaatttctctcatttggaatggctgtatttacccaatacctgcacacccat 1120

Qy   1726 tgyat 1730
          ||:||
Db   1121 tgtat 1125


RESULT  12
ABA90193
ID   ABA90193 standard; DNA; 465237 BP.
XX
AC   ABA90193;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Human oestrogen receptor alpha gene.
XX
KW   Human; oestrogen receptor alpha; ESR-alpha; ER; chromosome 6; Syne-2;
KW   synaptic nuclei expressed gene 2; haplotype; cytostatic; osteopathic;
KW   cardiant; vasotropic; gene therapy; vaccine; cancer; osteoporosis;
KW   cardiovascular disease; oestrogen receptor; ds.
XX
OS   Homo sapiens.
XX
```

PN WO200162969-A2.
XX
PD 30-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05358.
XX
PR 22-FEB-2000; 2000US-183756P.
PR 20-OCT-2000; 2000US-0692414.
PR 24-JAN-2001; 2001US-0768184.
XX
PA (PEKE) PE CORP NY.
XX
PI Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES;
XX
DR WPI; 2002-041152/05.
DR P-PSDB; AAG68251.
XX
PT Novel variant of estrogen receptor alpha polypeptide useful for
PT determining the biological activity of a protein for high throughput
PT screening and for raising antibodies that elicit an immune response in
PT host -
XX
PS Example; Fig 1 page 1-93; 333pp; English.
XX
CC The present invention describes an isolated peptide (I) consisting of an
CC amino acid sequence selected from: (a) the amino acid sequence of a
CC variant of the oestrogen receptor alpha (ESR-alpha) protein in AAG68251;
CC or (b) a fragment comprising at least 10 contiguous amino acids of the
CC protein in AAG68251. (I) has cytostatic, osteopathic, cardiant and
CC vasotropic activities, and can be used in gene therapy and vaccine
CC production. (I) is useful for identifying an agent that binds to (I), by
CC contacting (I) with an agent and assaying the contacted mixture to
CC determine whether a complex is formed with the agent bound to the
CC peptide. A polynucleotide (II), encoding (I), is useful in the
CC development of diagnostics and therapies for diseases and disorders
CC mediated/modulated by an oestrogen receptor (ER). (II) is also useful in
CC gene therapy for treating cancer, osteoporosis and cardiovascular
CC diseases. The human ESR-alpha gene is located on chromosome 6. The
CC present sequence represents the human ESR-alpha gene, which is given in
CC the exemplification of the present invention.
XX
SQ Sequence 465237 BP; 133988 A; 89578 C; 93946 G; 147721 T; 4 other;

Query Match 24.3%; Score 421; DB 24; Length 465237;
Best Local Similarity 46.5%; Pred. No. 2e-95;
Matches 477; Conservative 213; Mismatches 292; Indels 43; Gaps 7;

```
Qy    746 argarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnttyw 805
          |:||:||::: ||:||:||:: ||:||: |:|| || ||||| :| :| ||:|| ||:: 
Db 438788 aagaaatttctaagcagcaaaacattcaaaaggtgacttgggtgttgttaaaagcattct 438847

Qy    806 snttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnatha 865
          : ||:| ::| ||: |  |:||:||::: :: ||:||::| || || ||: |:|   |: 
Db 438848 gttttaaaagagaaac--agcataaaagttcagaaaatttgcagcctgatgatgcagtag 438905

Qy    866 araaraaraaycc-nttywsngarggnaarttyaarytngcngcngarathtgyathtgy 924
           :||:||:||:||  ||::  ||:|  ||:||:||::| || || ||:||:||:||: |: 
Db 438906 gaaagaaaaacccatttttttgaggagaaattcaagctggctgcagaaatttgcataagt 438965

Qy    925 aaygargarytnaaygtnaayccncargayaaygggngaraayathwsntggacntgycar 984
          ||: |:||::  ||:|| ||:|| ||:||:||:|| ||:||: |:::  |  | ||:|| 
Db 438966 aacaaggagccaaatgttaatccccaagacaatggggaaaatgtctccagagcatgtcat 439025

Qy    985 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg----naartggtt 1040
          :| :: ::   :||: :  |:  ::: :   |  ||:  || :| :|     ||:  ||| 
Db 439026 aggtcttcatggcagcccctcccatcacagacccggaagcctaggaggaaaaaaacagtt 439085

Qy   1041 ytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygt 1100
          :||:||  | |  || || :: :  ||:||:|| ||:|| :|  |::| || || ||: | 
Db 439086 ttgtgggccagtcccagggtccccatgctgtgtgcagcctaggaacttggtgccctgcat 439145

Qy   1101 nccngtnaaywsngcngtngc----------nwsngarggngcnws 1136
```

```
                |  |    : : ||  | ||                          :: ||:|  || :: 
Db 439146 ctcagctgctccagctattgctaaaaggggctgaggtaccacggtttcagaggttgcaag 439205

Qy   1137 nccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnat 1196
          || ||:|| ||||::|  | :: || || ||:|| ||  |  |  |:||::  :| || 
Db 439206 ccccaaaccttggcagctttcatgtggtgttgagcctgtgtgtacacagaagttaagaat 439265

Qy   1197 hgargtntgggarccnccnathmgnttycaraarathtayggnaayccntggatgccnmg 1256
          :||:|| |||||:|| ||  |::  ||:||:||:||:  :|| ||: | ||||| || :| 
Db 439266 tgaggtttgggaacctccacttatatttcagaagatacgtggaaatgcctggatacccag 439325

Qy   1257 ncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraa 1316
          ||:| :||:|| |  |  || ||  : :: |||:|  | :: || :| |  |  :|| 
Db 439326 gcaaacatttgctgcagaggtgggccctcatggagggcctctgctagggcaatgaggaa 439385

Qy   1317 rggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngc 1376
          :|| ||:|| || | |||:|| || ||::| || || :: || || |   : ::  | || 
Db 439386 gggaaatgtggggttggaacccccacacagagtccccactggggcactgcctagtggagc 439445

Qy   1377 ngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytnca 1436
          || :|  |  : || |  :: :: :| :  ||:||  | :| :: || ||::: :| || 
Db 439446 tgtaagaggaggaccactgtcctccagaccgcagaatagtagatccactgacagcttgca 439505

Qy   1437 rcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggna---- 1492
          ||:|| || | :||: : || ||:|| ||: |:||:|| || ||:|| |   |
Db 439506 ccatgtgcctggaaaagccacagacactcaacgccagcctgtgaaagcagtcagggttgg 439565

Qy   1493 ------tggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggna 1546
                |||  :  :  | || ||:||: |    |  || ||::|  | ||:||  | || | 
Db 439566 aggtggtggtggctataccctataaagccacaggggcagagctgcccaagactatgggaa 439625

Qy   1547 thtayytnytntncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyg 1606
          :||::| :| ||:   ||:||::| ||:|| :| ||:   || ||: | ||: |:||:| 
Db 439626 cctacctcttgcatcagcatgacctggatgtgagacattcagtcaaaggagat-attttg 439684

Qy   1607 gngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytn 1665
           || :|    ||:||:||:||    || ||:   || | : |||| || ||  || :| 
Db 439685 aagctttagaatttgactgccctggtggattttagacttgtgtgggccctgtaacccctt 439744

Qy   1666 tgyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytncay 1725
          ||:||:|| ||:||: |:|| ||:||  | || || ||:|| ||:||::|  |::  ||: 
Db 439745 tgttttggccaatttctcccatttggagctgctgtatttacccaatgcctaaacccgcat 439804

Qy   1726 tgyat 1730
          ||:||
Db 439805 tgtat 439809
```

RESULT 13
AAS91984
ID AAS91984 standard; cDNA; 1485 BP.
XX
AC AAS91984;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27788.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.

```
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27797.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 27788; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1485 BP; 442 A; 333 C; 373 G; 337 T; 0 other;


  Query Match             23.6%;  Score 409;  DB 23;  Length 1485;
  Best Local Similarity   46.5%;  Pred. No. 1.5e-94;
  Matches  458;  Conservative  199;  Mismatches  287;  Indels   40;  Gaps    6;

Qy     787 gtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnwsngaraayytn 846
           |  :  :|  | || ||::: ||:||::| ||:|| ||:||:||::  :  ||:||::|
Db     103 gcttctttgtatgcattcagttttataagggaagcagagcataaaaatttggaaaatttg 162

Qy     847 cayccngayaaygtnathaaraaraaraayccnttywsngarggnaarttyaarytngcn 906
           || || ||: |:|  ||: |:||:||:||:|| ||::: | :|  ||: |:||::
Db     163 cagcctgactatgcaatagaaaagaaaaacccattttctggggagaaactcaagcctatt 222

Qy     907 gcngarathtgyathtgyaaygargarytnaaygtnaaycncargayaayggngaraay 966
           || ||:||:||: |: |:| : |:||::  ||:|| ||:|| ||:||: |: | | :||:
Db     223 gcagaaatttgcctaagtagcaaggagcctaatgttaatccccaagaccatcggggaaat 282

Qy     967 athwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccn 1026
            |::   || | ||:||: | :: ::   :| : :  |:  ::: :  || || :| ||
Db     283 gtctacaggccatgtcagggaacttcatggctgcccctcccatcacaagcctgcaggccc 342

Qy    1027 mgnmg---naartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgn 1083
           :| :|    ||:|||||:  :||  | || || || :  :| ||: |:|| ||:|  :|
Db     343 aggagtaaaaagtggtttcatgggccaggcccagggttcctgtgccgtgtgcagcacagg 402

Qy    1084 gayytngtnccntgygtnccng----------------------------------tnaay 1110
           ||::| || || ||:|| || |                                   :
Db     403 gacttggtgccctgtgtcccagctgctccagctatggctgaaaggggccaaggtagaggt 462

Qy    1111 wsngcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngar 1170
           :: || || || :: ||:|| || :: || ||:|| ||||||::| ||  : |  || ||:
Db     463 tgggctgtggcttcagagggtgcaagacccaagccttggcagcttccacgtgatgttgag 522

Qy    1171 ccngtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraar 1230
           || || || ||  |:||::: :| | :||:|| ||||||:| |    ::| ||:||:||:
Db     523 cctgtgggtgcacagaagtcaagaactgaggtttgggaacctctgcctagatttcagaag 582

Qy    1231 athtayggnaayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntgg 1290
           || ||:|| ||: | |||||||| :| ||:||:||:|| |  || || ||  : :: |||
Db     583 atgtatggaaacgcctggatgcccaggcaaaagtttgctgcaggtgtgggaccctcatgg 642

Qy    1291 mgnacnwsngcnmgngtngtncaraarggnaaygtnggntgggar-ccnccncaymgngt 1349
           :| || :: || :| |  || ||:||:|  ||:|| || |  ||: || || ||::| ||
Db     643 agaacctctgctagggcagtgcagaaggaaaatgtggggtcagagccccccacacagagt 702

Qy    1350 nccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccncccnwsnwsnmgnytnca 1409
            || :: || || ||  : ::  | || || :| :|  : || |  :: :: :| :  ||
Db     703 ccctactggggcaccacctagtggagctgtgagaagggggccactgtcctccagacccca 762

Qy    1410 raarggnmgnwsna-cngaywsnytncarcaygtnccngaraarwsnacngayacncart 1468
           :|| || :| :: | |  |::: :| ||   :|  || | :||: : || || || ||:|
Db     763 gaatggtagatctacccaacagcttgcactgtgcacctggaaaagccacagaaactcaat 822

Qy    1469 gycarccngtnaargcngcnggnatggarw-sngtnccntayaaracngtngtngcngar 1527
           |:||: | || ||:||  |  | |  | :: : || || | :||: |    |  |  ||:
Db     823 gccagaccgtgaaagcaactaggagagggtgctgtaccctgcaaagccacaggggtggag 882

Qy    1528 ytnacnaaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggn 1587
           :|  | ||: |  | ||   : |: | :| ||:   ||:||::| ||:|| :| ||:||
Db     883 ctgcccaaggccatgggagcccacgtcttgcatcagcatgacctggatgttagacatgga 942

Qy    1588 gtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntay 1647
           || ||: | ||:||:||:|| || :| :  ||:||:||:|| || || ||::  || | :
Db     943 gtcaaaggagatcattttggggctttaaaatttgactgccctactggatttcagacttgc 1002

Qy    1648 atgggnccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyac 1706
            |||| || ||  || :| ||:||:|| ||:| : |:||  |:|| |  || || ||:||
Db    1003 gtgggccctgtaaccactttgttttggccaatatctcccagttggaatggctgtatttac 1062

Qy    1707 ncartgyytntayytncaytgyat 1730
           ||:| ::| ||::| ||:||:||
Db    1063 tcaatacctgtacctccactgcat 1086


RESULT  14
AAD05178
ID   AAD05178 standard; cDNA; 1278 BP.
XX
AC   AAD05178;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 14 cDNA clone HISBF60, SEQ ID NO:68.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             224..268
FT                   /*tag=  a
FT                   /product= "Human secreted protein"
FT                   /note= "CDS does not include start codon"
FT                   /partial
XX
PN   WO200134769-A2.
XX
PD   17-MAY-2001.
XX
```

```
PF   01-NOV-2000; 2000WO-US30040.
XX
PR   05-NOV-1999;   99US-0163580.
PR   30-JUN-2000; 2000US-0215130.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Wei P,  Baker KP,  Fiscella M;
XX
DR   WPI; 2001-308781/32.
DR   P-PSDB; AAE01289.
XX
PT   New isolated nucleic acid molecule encoding a human secreted protein is
PT   used in preventing, treating or ameliorating a medical condition -
XX
PS   Claim 1; Page 442; 519pp; English.
XX
CC   AAD05121-AAD05203 represent cDNAs corresponding to 24 human secreted
CC   protein genes, and AAE01232-AAE01311 represent the proteins they encode.
CC   AAE01312-AAE01340 represent human secreted protein variants or fragments.
CC   The secreted proteins and their genes are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   24 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 1278 BP; 344 A; 264 C; 358 G; 299 T; 13 other;
```

```
  Query Match             23.6%;  Score 408.8;  DB 22;  Length 1278;
  Best Local Similarity   45.8%;  Pred. No. 1.5e-94;
  Matches  458;  Conservative  205;  Mismatches  298;  Indels   40;  Gaps    6;

Qy     745 gargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntty 804
           ||:||:||::: ||:||:||::: ||:||:||:|| |  ||||| :| :| ||:|| ||:
Db     280 gaagaaatttctaagcagcaaagcattcaagaggtgatttgggtgctgttaaaagcattt 339

Qy     805 wsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnath 864
             ||:| ::| ||: |  |:||:||::  :  ||:||::| || || | : |:|    :
Db     340 cattttaaaagggaaac--agcataaaaatttggaaaatttgcagccagttgatgcagca 397

Qy     865 aaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtgy 924
            |: |:| :||:|  ||::  ||:|  ||:||:||::| || || ||:||:||:||: |:
Db     398 gaagagagaaacctgttttttgaggagaaattcaagctggctgcagaaatttgcataagt 457

Qy     925 aaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgycar 984
            |:  :||::| ||:|| ||:|  ||:||:|||:|| ||:||: |:::  || | ||:||
Db     458 tacagggagctgaatgttaatctccaagacaatggggaaaatgtctccagggcatgtcac 517

Qy     985 mgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaar---tggtty 1041
           :| :: ::   :||: :  |:  ::: :   |   |:   | :| :| ||:   |||||:
Db     518 aggtcttcatggcagcccctcccatcacagaccctgaagactaggaggaaaaaatggttt 577
```

```
Qy    1042 tgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgygtn 1101
           ||:||  | || || || :: :  ||:||:|| ||:|| :| ||::| || || ||: |
Db     578 tgtgggccaggcccagggtccccatgctgtgtgcagcctagggacttggtgccctgcatc 637

Qy    1102 ccngtnaaywsngcngtngc------------------------------nwsngarg 1129
           || |    : : ||  | ||                                :: ||:|
Db     638 ccagctgctccagccattgctaaaagggccgaggtacagctctgcccatggtttcagagg 697

Qy    1130 gngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarw 1189
           | || :: || ||:|| |||||::| || :: || || ||:|| |  || ||   :||::
Db     698 gtgcaagccccaaaccttggcagcttccatgtggtgttgagcctgcaggtgcatggaagt 757

Qy    1190 snmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntgga 1249
           : :| ||:||:|| |||||:|| ||   ::| ||:||:||:|| ||:|| | : | ||||
Db     758 caagaattgaggtttgggaacctccatctagatttcagaagatgtatggaatcacctgga 817

Qy    1250 tgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtng 1309
           |||  :| ||:||:||:||    || |  |   : :: |||:| || :: || :| |  |
Db     818 tgctcaggcaaaagtttgctccaggggcagagccttcatggagaacctctgctagggcag 877

Qy    1310 tncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnw 1369
           |   :||:|| ||:|| || | |||:|| |: ||::|  | || :: || || ||  : :
Db     878 tg-tgaagggaaatgtggggttggagccccyacacagaatccctactggggcaccaccta 936

Qy    1370 snmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayw 1429
           :  | || || :| :| :: || || :: :  :| :  ||:|| || :| :: || ||::
Db     937 gtggagctgtgcgaagaaggccaccgtccttcagaccccagaatggtrgatccaccgaca 996

Qy    1430 snytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcng-cn 1488
           : :|  | | :|  || | :||: : || ||: |  |: |:||:|| || ||:|| | |
Db     997 gcttgtaccgwgcacctggaaaagccacagaccctaaacgccagcccgtgaaagcagcca 1056

Qy    1489 ggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnath 1548
           ||  ||| :::  | || | : |: |    |  || ||::|  | ||:||    || | :
Db     1057 ggagtggggactataccctgtgaagccacaggggcagagctgcccaagactaagggaacc 1116

Qy    1549 tayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggn 1608
           ||::| :| ||:  :   ||::|  |:|| :| ||:|| || | : | ||:||:|:|
Db     1117 tacctcttgcatcattgngacctgnatgtgagacatggagtcagaggagatcattttgna 1176

Qy    1609 gcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytntgy 1668
            |  |    ||:||:||:|   | || ||: | || | :|||||  |  |   | :  | :
Db     1177 acgntataatttgactgcctcgctggattttggacttgcatggggcttgnaacgcctttt 1236

Qy    1669 ttyg-gncarttyttyccnttyggnacngcngtnttyacnc 1708
              |:| | ||:||: |:|| ||:|| |  || || ||:|| |
Db     1237 gttgtggcagtttctcccatttggaatggctgtatttaccc 1277
```

```
RESULT  15
AAS87262/c
ID   AAS87262 standard; cDNA; 1580 BP.
XX
AC   AAS87262;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #23066.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
```

XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG23075.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 23066; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1580 BP; 357 A; 412 C; 328 G; 483 T; 0 other;

Query Match 23.1%; Score 400.4; DB 23; Length 1580;
Best Local Similarity 42.4%; Pred. No. 2.6e-92;
Matches 466; Conservative 215; Mismatches 383; Indels 34; Gaps 4;

```
Qy     661 aaytaytaywsnwsnttygaygtngarytncaycaytgyaargtnaayathtggytnatg 720
           ||:| : |::      ||:   |   | |    : |: | || ||:  :  |:
Db    1107 AATTCCGATAACGAACAAGACTGTGGCATGCTAACTAGTTACGTGAACCCCGAGCAGTCA 1048

Qy     721 mgnaarcarathytngcnaayaargargarathwsnaarcarcarwsnathcargargtn 780
           |  | ::|     |   : |:| :  ::  |  || | :::  |      |
Db    1047 GCATCCCCCAACTTCTTAGAGGGACAAGTGGCATTCAGCCACCTGAGATTGAGCAATAAC 988

Qy     781 acntgggtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnwsngar 840
           |  || |  :|      :|   |::  ||:| ::| ||:|| ||:||:||::  :  | :
Db     987 AGATGTGGTTTGGAATTGGAACTTATATTTAAAAGGGAAGCAGAGCATAAAAAAATTGGA 928

Qy     841 aayytncayccngayaaygtnathaaraaraaraayccnttywsngarggnaarttyaar 900
           ||::| || |  ||:| :|  ||: |:||:||:||:|| ||::: | :|  ||:||:||:
Db     927 AATCTGCAGCATGACAGTGCAATAGAAAAGAAAAACCCATTTTCTGGGGAGAAATTCAAG 868

Qy     901 ytngcngcngarathtgyathtgyaaygargarytnaaygtnaayccncargayaayggn 960
           :| |  || ||:||:||:||:  :||: |:||::  ||:|| ||:|  ||:||:||:||
Db     867 CTGGTTGCAGAAATTTGCATAAATAACAAGGAGCCAAATGTTAATCACCAAGACAACGGG 808

Qy     961 garaayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntgg 1020
           |:||: |::   || |  |:||::|  : :     ||: :  |:  ::: :  ||   |
Db     807 AAAAATGTTTTCAGGGCACGTCAGAGACCCTTATGCCAGCCCCTCCCATCGCAGGCCCAG 748

Qy    1021 mgnccnmgnmgnaa---rtggttytgyggnacnggnccnggnwsnytntgytgygtncar 1077
           :| || :| :| ||   :|||||:||:||    || || ||  : :  ||:| :|| ||:
Db     747 AGGCCTAGGAGGAAAATATGGTTTTGTGGGCTGGGCCCAGGACCTTGCTGCTTTGTGCAG 688

Qy    1078 ccnmgngayytngtnccntgygtnccngtnaaywsng---------------------- 1114
```

```
           | :  ||::| || || ||: | ||        :
Db     687 TCTCAGGACTTGGTGCCCTGCATCCCATCTGTGGCTAAAAGGGGCCAATATACAGCTGAA 628

Qy    1115 -cngtngcnwsngargggngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 1173
            |  | || :: ||:|| |  :: || ||:|  |||  ::|  | :: || || | :||
Db     627 ACCATTGCTTCAGAGGGTGTAAGCCCCAAGCATTGGAGGCTTACATGTGGTGTTGGACCT 568

Qy    1174 gtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarath 1233
           || ||  |  |:||::: :| | :||:|| ||||:|| |    ::| ||:||:| :||
Db     567 GTGGGTACACAGAAGTCAAGAACTGAGGTTTGGGAACCTCTGCCTAGATTTCAGAGGATG 508

Qy    1234 tayggnaayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgn 1293
           ||:|| ||: | |||||  | :| ||:| :| :|| |  || || |   : :: |||:|
Db     507 TATGGAAACACCTGGATATCCAGGCAGAGGTGTGCTGCAGGGGTAGAGCCCTCATGGAGA 448

Qy    1294 acnwsngcnmgngtngtncaraarggnaaygtnggntgggarccncccncaymgngtnccn 1353
           || :: || :| |   | ||:||:|| ||:|| || | |||:|| || ||::| || ||
Db     447 ACCTCTGCTAGGGCAATGCAGAAGGGAAATGTGGGGTTGGAGCCTCCACACAGAGTCCCC 388

Qy    1354 wsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraar 1413
           :: ||       :: ::  | || || :| :|  : || || :: :: :| :  ||:||
Db     387 ACTGG------GACTAGTGGAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAAT 334

Qy    1414 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgycar 1473
           |  :|  : || ||::: :|  | | :|| || | :||: : ||  |:|| ||:||:||
Db     333 GATAGACCAACTGACAGCTTGGATCGTGTACCTGGAAAAGCCACAAACACTCAATGCCAT 274

Qy    1474 ccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacn 1533
           || || ||:|| ||  | | || : : || || | :||: |    |  || ||::|  |
Db     273 CCTGTGAAAGCAGCCAGGAGGGGAGCTGTACCCTGCAAAGCCACAGGGGCAGAGCTTCCC 214

Qy    1534 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1593
           ||: | || |  | : |::| :| ||:   | :| ::| ||:|| :| | :|| || ||:
Db     213 AAGGCTGTAGAAACCCACCTCTTACATCAGCGTGTCCTGGATGTGAGACCTGGAGTCAAA 154

Qy    1594 mgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggn 1653
           | ||:||:||:|| |  :| :| ||:| :||:|| || || ||::| |  | :|||||
Db     153 GGAGATCATTTTGGAGTTTTAAGATTTGGCTGCCCCACTGGATTTAGGAGTTGCATGGGG 94

Qy    1654 ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg 1712
           || ||  || :| ||:||:|| ||:||: |:|| | :|| |     || ||:|  ||:||
Db      93 CCTGTAGCCCCTTTGTTTTGGCCAATTTCTCCCATCTGGAATGAGTGTGTTTATCCAATG 34

Qy    1713 yytntayytncaytgyat 1730
           ::  ||::  ||:| :||
Db      33 CCCGTACCCCCATTATAT 16
```

Search completed: July 8, 2002, 16:25:55
Job time: 12728 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 16:20:40 ; Search time 104.62 Seconds
(without alignments)
4064.152 Million cell updates/sec

Title: US-09-997-610-7
Perfect score: 1731
Sequence: 1 atgytnathgargcntaygg.....gyytntayytncaytgyatg 1731

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	299.2	17.3	49136	4	US-09-422-869-1	Sequence 1, Appli
c	2	177.6	10.3	14855	2	US-08-687-080-59	Sequence 59, Appl
	3	164.6	9.5	620	4	US-08-991-789A-29	Sequence 29, Appl
	4	164.6	9.5	620	4	US-09-062-451-29	Sequence 29, Appl
	5	140.6	8.1	152331	3	US-09-128-155-16	Sequence 16, Appl
	6	135.8	7.8	14855	2	US-08-687-080-59	Sequence 59, Appl
c	7	67	3.9	7218	1	US-08-232-463-14	Sequence 14, Appl
	8	59	3.4	2067	4	US-09-106-194-11	Sequence 11, Appl
	9	51.8	3.0	1313	2	US-08-463-911-6	Sequence 6, Appli
	10	51.8	3.0	4517	4	US-09-140-804-9	Sequence 9, Appli
c	11	50.6	2.9	1155	3	US-09-053-866-3	Sequence 3, Appli
	12	48.8	2.8	2082	4	US-09-440-325A-2	Sequence 2, Appli
	13	48.8	2.8	2088	4	US-09-351-414-3	Sequence 3, Appli
	14	48.6	2.8	289	4	US-09-007-005-17	Sequence 17, Appl
	15	48.6	2.8	289	4	US-09-244-796-17	Sequence 17, Appl
	16	48.6	2.8	1107	3	US-09-188-930-217	Sequence 217, App
	17	48.6	2.8	1276	2	US-08-463-911-1	Sequence 1, Appli
	18	48.2	2.8	2277	1	US-08-676-967-2	Sequence 2, Appli
	19	48.2	2.8	2277	1	US-08-676-974-2	Sequence 2, Appli
	20	48.2	2.8	2277	2	US-09-098-487-2	Sequence 2, Appli
c	21	48	2.8	2277	1	US-08-676-967-2	Sequence 2, Appli
c	22	48	2.8	2277	1	US-08-676-974-2	Sequence 2, Appli
c	23	48	2.8	2277	2	US-09-098-487-2	Sequence 2, Appli
	24	44.8	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	25	44.4	2.6	1212	2	US-09-092-770-18	Sequence 18, Appl
c	26	44.4	2.6	1212	4	US-09-222-851-18	Sequence 18, Appl
	27	43.4	2.5	5181	1	US-08-257-073-10	Sequence 10, Appl
	28	43.2	2.5	729	4	US-09-140-804-10	Sequence 10, Appl
c	29	43	2.5	1215	2	US-09-092-770-8	Sequence 8, Appli
c	30	43	2.5	1215	4	US-09-222-851-8	Sequence 8, Appli
	31	42.4	2.4	1015	3	US-09-188-930-30	Sequence 30, Appl
c	32	42	2.4	2067	4	US-09-106-194-11	Sequence 11, Appl
c	33	42	2.4	35060	3	US-08-814-095-7	Sequence 7, Appli
	34	41.8	2.4	1001	3	US-09-188-930-218	Sequence 218, App
	35	41.8	2.4	2265	3	US-09-369-618-3	Sequence 3, Appli
	36	41.8	2.4	2265	3	US-09-369-617-3	Sequence 3, Appli
	37	41.6	2.4	510	2	US-08-934-959-7	Sequence 7, Appli
c	38	41	2.4	597	4	US-09-528-760A-3	Sequence 3, Appli
c	39	40.8	2.4	723	3	US-08-911-423-5	Sequence 5, Appli
	40	40.8	2.4	1659	1	US-08-231-729B-2	Sequence 2, Appli
	41	40.6	2.3	2949	4	US-09-412-554A-3	Sequence 3, Appli
	42	40.2	2.3	500	3	US-09-141-000-2	Sequence 2, Appli
	43	40.2	2.3	1659	1	US-08-231-729B-1	Sequence 1, Appli
c	44	40	2.3	2088	4	US-09-351-414-3	Sequence 3, Appli
c	45	39.8	2.3	2265	3	US-09-369-618-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-422-869-1
; Sequence 1, Application US/09422869
; Patent No. 6235481
; GENERAL INFORMATION:
; APPLICANT: POLONSKY, KENNETH S.
; APPLICANT: HORIKAWA, YUKIO
; APPLICANT: ODA, NAOHISA
; APPLICANT: COX, NANCY J.
; APPLICANT: SREENAN, SEAMUS
; APPLICANT: ZHOU, YUN-PING
; APPLICANT: OTANI, KENICHI
; APPLICANT: HANIS, CRAIG L.
; APPLICANT: BELL, GRAEME I.
; TITLE OF INVENTION: METHODS OF TREATMENT OF TYPE 2 DIABETES
; FILE REFERENCE: ARCD:307
; CURRENT APPLICATION NUMBER: US/09/422,869
; CURRENT FILING DATE: 1999-10-21
; EARLIER APPLICATION NUMBER: 60/134,175
; EARLIER FILING DATE: 1999-05-13
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 49136
; TYPE: DNA
; ORGANISM: Human
US-09-422-869-1

Query Match 17.3%; Score 299.2; DB 4; Length 49136;
Best Local Similarity 41.1%; Pred. No. 9.8e-74;
Matches 388; Conservative 186; Mismatches 307; Indels 62; Gaps 5;

```
Qy    798 rgcnttywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaa 857
          :|   |::: ||:| ::| ||:|| ||:||:||::: :  ||:||::| ||:|| ||: |
Db  47384 ggaactcacatttaaaagggaagcagagcataaaagtttagaaaatttgcatcctgacta 47443

Qy    858 ygtnathaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtg 917
          :|  ||: |:||:||:||: | ,||::: ||:|  ||:||:||::  || || ||:||:||
Db  47444 tgggatagaaaagaaaaactcattttctgaggagaaattcaagccagctgcagaaatttg 47503

Qy    918 yathtgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggac 977
          :||: |:||: |:||::   |:|| ||:     :||:||:|  | :||: |:::  || |
Db  47504 cataagtaactaggagccacatgttaatagcatagacaatgagggaaatgtctccagggc 47563

Qy    978 ntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naa 1034
          ||:||::| :: :: ||:||: :   :  ::: :  || |||:| |  :| :|    ||
Db  47564 atgtcagaggtcttcacagcaacccccacccatcacaggcctggaggcttaggaggaaaaa 47623

Qy   1035 rtggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtncc 1094
```

```
          :|||||:||: |    || || ||  : :  || | :|| ||: | :  ||::| || ||
Db  47624 atggttttgtcgttggggcccagggccttgctggtttgtgcagtctcaggacttggtgcc 47683

Qy   1095 ntgygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggcaryt 1154
            : | |           | |  |   :  |: | |        | : | | |   :|
Db  47684 ccacatcc-----------cagcagtggctaaaaggggccaatgtacagcttagacctt 47731

Qy   1155 nccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarccncc 1214
            | :: |  ||  |: |                                        ||
Db  47732 tgcttcagaggtgcaagc----------------------------------------cc 47751

Qy   1215 nathmgnttycaraarathtayggnaaycentggatgccnmgncaraarttygcngtngg 1274
           |  :  |    :   | :| :||   :           :  ||:||:||:|| |
Db  47752 caagccttggtggcttacatgtggtgttgggcctgcagatacacagaagtttgctgcact 47811

Qy   1275 ngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntggga 1334
           || |   : :: || :| || :: || :| |  ||  |:||: |  |:|| || | |||
Db  47812 ggtggaaccctcatgtagaacctctgctagggcagtgtagaagtgatatgtggggttgga 47871

Qy   1335 rccnccncaymgngtnccnwsnggngcnc------cnwsnwsnmgngcngtnmgnmgnws 1388
          :|| || ||::   | || :: || || |      |  : ::  |  | || :| :| ::
Db  47872 gccccccacacaatccccactggggcactgcctactgctactggaactgtgagaagaag 47931

Qy   1389 nccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccnga 1448
           || || :: :: :| :  ||:|| || :| :: || ||: : :|  | ||:|  || |
Db  47932 gccaccatcctccagaccccagaatggtagatccactgatggcttgaaccatgcacctgg 47991

Qy   1449 raarwsnacngayacncartgycarccngtnaargcngcngggnatggarwsngtnccnta 1508
          :||: : || ||:|| ||:  :|||:|| || ||:|| || || | |||: : || || |
Db  47992 aaaagccacagacactcaacaccagcctgtgaaggcagctggaagggaggctgtaccctg 48051

Qy   1509 yaaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycayga 1568
          :||:||    |  || ||::|  | ||:    | ||   : |:::| :| ||:   | :||
Db  48052 caaaacaacagaggcagagctgcccaaggtcatgggagcccacctcttgcatgagcctga 48111

Qy   1569 yytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgycc 1628
          ::|  |:|| :| ||:|| || ||: | ||:||:||:|| || :| :| ||:||:||:||
Db  48112 cttgaatgtgagacatggagtcaaaggagatcattttggagctttaagttttgactgccc 48171

Qy   1629 nacnggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnt 1687
            | || ||::| || | :|||| || ||  || :|   :||:|| ||:||: |:|| |
Db  48172 acctggatttcggacttgcatggggcctgtggccccttcattttggccaatttatcccat 48231

Qy   1688 tyggnacngcngtnttyacncartgyytntayytncaytgyat 1730
           |:|| |  |   | ||:|| ||:||::| ||::  ||:| :||
Db  48232 ttggaatgggtatatttacccaatgcctgtacccccattctat 48274


RESULT   2
US-08-687-080-59/c
; Sequence 59, Application US/08687080
; Patent No. 5965427
;  GENERAL INFORMATION:
;    APPLICANT:  Gregory Dolganov
;    TITLE OF INVENTION:  Human RAD50 Gene and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  175
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/687,080
;      FILING DATE:  17-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/592,126
;      FILING DATE:  26-JAN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  4600-0111.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14855 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  5' END OF INTRON 2 OF RAD50 GENOMIC
;      INDIVIDUAL ISOLATE:  SEQUENCE
US-08-687-080-59


  Query Match             10.3%;  Score 177.6;  DB 2;  Length 14855;
  Best Local Similarity   44.4%;  Pred. No. 2.2e-39;
  Matches  221;  Conservative   90;  Mismatches  172;  Indels   15;  Gaps    4;

Qy   1244 cntggatgccnmgncaraarttygcngtnggngtnggnwsnwsnt9gmgnacnwsngcnm 1303
          | |||||||| :| ||:| :||:|| |  || || ||  : :  |||:  || :: || :
Db  14854 CCTGGATGCCCAGGCAGATGTTTGCTGCAGGGGTGGGGCCCTTATGGAAAACCTCTGCTA 14795

Qy   1304 gngtngtncaraarggnaaygtnggngtgggarccncncaymgngtnccnwsnggngcnc 1363
          | |   |   :||:|| ||:|| || | | |:|| |          | |  ::  | |  |
Db  14794 GGGCAATATGGAAGGGAAATGTGGGGTTGAAACCCCAC--AGAGTTCCTATGGAGGGGAC 14737

Qy   1364 cnwsnwsnmgngcngtnmgnmgnwsnccncnwsnwsnmgnyt----------ncaraar 1413
            : ::  | || || :| :|    || |  :: :: :| :|           ||:||
Db  14736 TGCCTAGTGGAGCTGTGAGAAGACAGCCACTGTCCTCCAGACTGGTAGATCCCCCAGAAT 14677

Qy   1414 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgycar 1473
            :| :: || ||::: :| ||   :|  || | :||:::  | | :|| ||:  :||:
Db  14676 AATAGATCCACTGACAGCTTGCACTGTGCACCTGGAAAAACTGCAGGCACTCAACACCAG 14617

Qy   1474 ccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacn 1533
          || || ||: | ||  | | |||: :  | || | :||: |    |  | ||::|  |
Db  14616 CCTGTGAAAACAGCCAGGAAGGAGGCTATACCCTGCAAAGCC--AGAAGTGGAGCTGCCC 14559

Qy   1534 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1593
          ||: |  | |    : |::| :| ||:     :||::| ||:|| :| ||:|| || ||:
Db  14558 AAGGCCATGGAAGCCCACCTCTTGCATCAGAGTGACCTGGATGTGAGACATGGAGTCAAA 14499

Qy   1594 mgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggn 1653
            | ||:||:| :|| || :| :| | :  :||:|| || |  ||::| || | :| |||
Db  14498 GGAGATCATTCTGGAGCTTTAAGATACACCTGCCCCACTGAATTTCGGACTTGCACGGGG 14439

Qy   1654 ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg 1712
          || ||  || :| ||:||:|| ||:||: |:|| ||:|| |  || || ||: | ||:||
Db  14438 CCTGTAGCCCCTTTGTTTTGGCCAATTTCTCCCATTTGGAATGGCTGTATTTGCCCAATG 14379

Qy   1713 yytntayytncaytgyat 1730
           ::| ||::  ||:||:||
Db  14378 CCTGTATCCCCATTGTAT 14361


RESULT   3
US-08-991-789A-29
; Sequence 29, Application US/08991789A
; Patent No. 6225054
;  GENERAL INFORMATION:
;       APPLICANT: Frudakis, Tony N.
;                  Smith, John M.
;                  Reed, Steven G.
```

```
;       TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
;                           TREATMENT AND DIAGNOSIS OF BREAST CANCER
;       NUMBER OF SEQUENCES: 292
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Seed IP Law Group
;            STREET: 70I Fifth Avenue, Suite 6300
;            CITY: SeattIe
;            STATE: Washington
;            COUNTRY: USA
;            ZIP: 98104-7092
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #I.0, Version #I.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/08/991,789A
;            FILING DATE: Il-Dec-I997
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Potter, Jane E. R.
;            REGISTRATION NUMBER: 33,332
;            REFERENCE/DOCKET NUMBER: 21012I.419C3
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206) 622-4900
;            TELEFAX: (206) 682-603I
;  INFORMATION FOR SEQ ID NO: 29:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 620 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 29:
US-08-99I-789A-29


  Query Match             9.5%;  Score 164.6;  DB 4;  Length 620;
  Best Local Similarity   40.7%;  Pred. No. 4.3e-37;
  Matches  250;  Conservative  I05;  Mismatches  243;  Indels   I7;  Gaps    6;

Qy   II26 garggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaar II85
          ||:||  | :: || ||:|| || ||::| || :  || ||  |:|| || || ||  |:
Db       3 GAGGGTACCAGCCCCAAGCCTTGACAACTTCCATAGGGTGTCAAGCCTGTGGGTGCACAG 62

Qy   1I86 aarwsnmgnathgargtntgggarccnccnathmgnttycaraarathtaygg------- I238
          ||::: :  ||:||: | ||||| || |    ::| ||:||:| :||:||  |
Db      63 AAGTCAAAAATTGAGTTTTGGGATCCTCAGCCTAGATTTCAGAGGATATAAAGAAACACC I22

Qy   1239 naaycntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnws I298
           ||: | | |||    :| ||:||:||: |    ||  |       ::  ||:  || ::
Db    123 TAACACCTAGATATTCAGACAAAAGTTTACTACAGGGATGAAGCTTTCACGGAAAACCTC I82

Qy   1299 ngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsngg I358
            | :|    || ||:||: | ||:|| || | |||:|| ||  |::|  | || ::  |
Db    183 TACTAGGAAAGTACAGAAGAGAAATGTGGGTTTGGAGCCCCCAAACAGAATCCCCTCTAG 242

Qy   I359 ngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmg I4I8
            | |   : :      | || :| :| :: || |  :: :: :| :  ||:|| |  :|
Db    243 AACACTGCCTAATGAAACTGTGAGAAGATGGCCACTGTCATCCAGACACCAGAATGATAG 302

Qy   I419 nwsnacngaywsnytncarcayg-tnccngaraarwsnacngayacncartgycarcc-- I475
            : ||  | :  :|    ||:  | ||   :||: : || ||:|| ||:||:||:||
Db    303 ACCCACCAAAAACTTATGCCATATTGCCTATAAAACCTACAGACACTCAATGCCAGCCCC 362

Qy   I476 -ngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacna 1534
             |  ||:    | |  | | |::: ||||| ||:||  |    |  || ||::|  |.
Db    363 ATGAAAAAAAAACTGAGAAGAAGACTGTNCCCTACAATGCCACCGGAGCAGAACTGCCCC 422

Qy   1535 aracngtnggnathtayytnytncaytgycaygayytngayg-tnmgncayggngtnaar I593
           |: |  | |    : | :| :|  |:     :||::| ||:| | :| ||:||  |   :
Db    423 AGGCCATGGAAGCACAGCTCTTATATCAATGTGACCTGGATGTTGAGACATGGAATCCNA 482

Qy   1594 mgngaycayttyggngcnytnmgntt----ygaytgyccnacnggnttymgnacntayat 1649
```

```
           |  |:| :||:  | | :  :   |    :||:||:|| |     ||: | || || ||
Db    483 NGAAATCNTTTTAANACTTCCACGGTTNAATGACTGCCCTATTANATTCNGAACTTANAT 542

Qy   I650 g-ggnccngtnccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacnc I708
            | || ||  | :  |    |: | | :||:  :|  ||:|| |  ||  | ||:   |
Db    543 CCNGGCCTGTGACCTCTTTGCTTTGGCCATTCCCCCTTTTTGGAATGGCTNTTTTTTTCC 602

Qy   I709 artgyytntayytnc I723
           :||::| | ::  |
Db    603 CATGCCTGTNCCCTC 617


RESULT   4
US-09-062-45I-29
; Sequence 29, Application US/09062451
; Patent No. 6344550
;  GENERAL INFORMATION:
;    APPLICANT:  Frudakis, Tony N.
;    APPLICANT:  Smith, John M.
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE
;    TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF BREAST CANCER
;    NUMBER OF SEQUENCES:  297
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98I04-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  FIoppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #I.0, Version #I.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/062,45I
;      FILING DATE:  04-APR-I997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  3I,392
;      REFERENCE/DOCKET NUMBER:  2I0I2I.4I9C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-603I
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  620 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-062-45I-29


  Query Match             9.5%;  Score I64.6;  DB 4;  Length 620;
  Best Local Similarity   40.7%;  Pred. No. 4.3e-37;
  Matches  250;  Conservative  I05;  Mismatches  243;  Indels   I7;  Gaps    6;

Qy   II26 garggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaar II85
          ||:||  | :: || ||:|| || ||::| || :  || ||  |:|| || || ||  |:
Db       3 GAGGGTACCAGCCCCAAGCCTTGACAACTTCCATAGGGTGTCAAGCCTGTGGGTGCACAG 62

Qy   II86 aarwsnmgnathgargtntgggarccnccnathmgnttycaraarathtaygg------- I238
          ||::: :  ||:||: | ||||| || |    ::| ||:||:| :||:||  |
Db      63 AAGTCAAAAATTGAGTTTTGGGATCCTCAGCCTAGATTTCAGAGGATATAAAGAAACACC I22

Qy   I239 naaycntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnws I298
           ||: | | |||    :| ||:||:||: |    ||  |       ::  ||:  || ::
Db    I23 TAACACCTAGATATTCAGACAAAAGTTTACTACAGGGATGAAGCTTTCACGGAAAACCTC I82

Qy   1299 ngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsngg I358
            | :|    || ||:||: | ||:|| || | |||:|| ||  |::|  | || ::  |
```

```
Db    183 TACTAGGAAAGTACAGAAGAGAAATGTGGGTTTGGAGCCCCCAAACAGAATCCCCTCTAG 242

Qy   1359 ngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmg 1418
          | |   : :      | || :| :| :: || |  :: :: :| :  ||:|| |  :|
Db    243 AACACTGCCTAATGAAACTGTGAGAAGATGGCCACTGTCATCCAGACACCAGAATGATAG 302

Qy   1419 nwsnacngaywsnytncarcayg-tnccngaraarwsnacngayacncartgycarcc-- 1475
          : ||  | :  :|    ||:  | ||   :||: : || ||:|| ||:||:||:||
Db    303 ACCCACCAAAAACTTATGCCATATTGCCTATAAAACCTACAGACACTCAATGCCAGCCCC 362

Qy   1476 -ngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacna 1534
           |  ||:    | |  | | |::: ||||| ||:||  |    |  || ||::|  |
Db    363 ATGAAAAAAAAACTGAGAAGAAGACTGTNCCCTACAATGCCACCGGAGCAGAACTGCCCC 422

Qy   1535 aracngtnggnathtayytnytncaytgycaygayytngayg-tnmgncayggngtnaar 1593
          |: |  | |    : | :| :|  |:     :||::| ||:| | :| ||:||  |   :
Db    423 AGGCCATGGAAGCACAGCTCTTATATCAATGTGACCTGGATGTTGAGACATGGAATCCNA 482

Qy   1594 mgngaycayttyggngcnytnmgntt----ygaytgyccnacnggnttymgnacntayat 1649
           |  |:| :||:  | | :  :   |    :||:||:|| |    ||: | || || ||
Db    483 NGAAATCNTTTTAANACTTCCACGGTTNAATGACTGCCCTATTANATTCNGAACTTANAT 542

Qy   1650 g-ggnccngtnccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacnc 1708
           | || ||  | :  |   |: | | :||:  :|  ||:|| |  ||  | ||:   |
Db    543 CCNGGCCTGTGACCTCTTTGCTTTGGCCATTCCCCCTTTTTGGAATGGCTNTTTTTTTCC 602

Qy   1709 artgyytntayytnc 1723
           :||::| | ::  |
Db    603 CATGCCTGTNCCCTC 617


RESULT   5
US-09-128-155-16
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
;  APPLICANT: Pan, Yang
;  TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 09404/052001
;  CURRENT APPLICATION NUMBER: US/09/128,155
;  CURRENT FILING DATE: 1998-08-03
;  EARLIER APPLICATION NUMBER: US 60/091,650
;  EARLIER FILING DATE: 1998-07-02
;  EARLIER APPLICATION NUMBER: US 60/054,646
;  EARLIER FILING DATE: 1997-08-04
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 16
;  LENGTH: 152331
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(152331)
;  OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16


  Query Match             8.1%;  Score 140.6;  DB 3;  Length 152331;
  Best Local Similarity   32.8%;  Pred. No. 1e-27;
  Matches  324;  Conservative  168;  Mismatches  426;  Indels   71;  Gaps    6;

Qy    771 hcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngarcayaa 830
          :||:| :   |     |  :|     :|    ::  ||:| ::  ||:|  ||:  :||
Db  71225 tcaggggtaaagaaatgacttaaagttggaacctatgtttaaaatggaagtagagtctaa 71284

Qy    831 rwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccnttywsngargg 890
          ::  :  ||:||::| || || | :  :||   : |:||: |:  :   |   : | :|
Db  71285 aaatttggaaaatttgcagcctggccttgtggcagagaaagaatccaagtaggctgcaga 71344

Qy    891 naarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaayccnca 950
            |:|    ::| |     ||:||: |:||   :||   :||::  | :|  ||:   ||

Db  71345 gcaatcattgctag----agagattagcatgactaaaagggagccaagtgctaatattca 71400

Qy    951 rgayaayggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnathaarws 1010
          :||:||:|   |:||   ::   || | | :||::| :: :   |:||: :  |:  :::
Db  71401 agacaatgttaaaaaggccttgagggcatttcagagatctatgaagcagcccctcccatc 71460

Qy   1011 nytngcntggmg-------nccnmgnmgnaartggttytgyggnacnggnccnggnwsny 1063
           :  |    |:|        |          |: ||||:| :||  |    || ||  : :
Db  71461 acaggtgcagaggtttggtgcactaggcccagaggttttatgggccannggcagggccac 71520

Qy   1064 tntgytgygtncarccnmgngayytngtnccntgygtnccngtnaa-------------- 1109
            ||:| :|  ||:|   | ||:   |   |  |: | |  |  |
Db  71521 actgctatgcacagctttgggacactgctgcccgcatccaggccactctgctctggctcc 71580

Qy   1110 ------------------------------ywsngcngtngcnwsngarggngcnwsncc 1139
                                          :::  |    ||  : ||:||  | :: |
Db  71581 acccttggctcaaacgggccaagatagagcttggaccactgctcccgagggcacaagcca 71640

Qy   1140 naarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnath-- 1197
           ||:|| |||  ::| || :: || ||  |:|| |  || ||  |:||  : :|  |
Db  71641 taagccttggtggtttccatgtggtgttaagcctgcaggtgcccagaatgcaagattgag 71700

Qy   1198 gargtntgggarccnccnathmgnttycaraarathtayggnaaycentggatgccnmgn 1257
           | :|  |||| :|  ||   ::  ||:||:| :|| | :   || || | | | || :|
Db  71701 ggagcttgggcacttccacctaaatttcagaggatgtgtcagaaaccctaggttcccagg 71760

Qy   1258 caraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraar 1317
           ||:||:  :|    || |  |   : :    |:| || ::  | :| |   | | :||:
Db  71761 cagaagcatgatacaggggcagagcccttgcagagaacctctactagggcaatgccaaag 71820

Qy   1318 ggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcn 1377
           |  ||:|| || | |||: |  | ||::  || || :: || || |   :  :     |
Db  71821 gaaaatgtggggttggagtcctcacacatggtccccactggggcactacctggtgatact 71880

Qy   1378 gtnmgnmgnwsnccncnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncar 1437
           ||  | :   : | |   | : :: :| :  ||:|| || :| :: || | ::: :  ||
Db  71881 gtgggaatggggctgctgccctccagaccccagaatggtagatgcactggcagctggcac 71940

Qy   1438 caygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggnatggar 1497
           | :|  || | :||: :  | | :|| ||:  :||:||  | | : | ||      ||
Db  71941 cctgagcctggaaaagctgcaggcactcaactccaacccatgagatcagccacatgggct 72000

Qy   1498 wsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtayytnytn 1557
           ::  | ||    ||: |    |  || | ::|  | ||: |  | ||   :||::  :|
Db  72001 ac--tcccagggaagcccacagaggcagggctgtctaaggccttgggagcctacccctg 72058

Qy   1558 caytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgn 1617
           |:   | :|     |:|  :   |            ||: |: |:|  || :| :|
Db  72059 aaccagcttgcaggacatggaatcaa-----------agattatgttgcagctttaagg 72106

Qy   1618 ttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytntgyttyggncar 1677
           |: |:  :    | |     :  |    : || |     |    :| | :||:
Db  72107 cttaatgttttccctgtcaatttcaggcttgtgtgggacctgttgctttttttttttttt 72166

Qy   1678 ttyttyccnttyggnacngcngtnttyac 1706
           ||:||:   || |  || |  || |  ||
Db  72167 ttttttttttggtcacaggtgtttgaac 72195


RESULT   6
US-08-687-080-59
; Sequence 59, Application US/08687080
; Patent No. 5965427
;  GENERAL INFORMATION:
;    APPLICANT:  Gregory Dolganov
;    TITLE OF INVENTION:  Human RAD50 Gene and Methods of Use Thereof
;    NUMBER OF SEQUENCES:  175
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
```

```
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/687,080
;      FILING DATE:  17-JUL-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/592,126
;      FILING DATE:  26-JAN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  4600-0111.30
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  59:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14855 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  5' END OF INTRON 2 OF RAD50 GENOMIC
;      INDIVIDUAL ISOLATE:  SEQUENCE
US-08-687-080-59


  Query Match             7.8%;  Score 135.8;  DB 2;  Length 14855;
  Best Local Similarity   39.5%;  Pred. No. 2.2e-27;
  Matches  191;  Conservative  101;  Mismatches  160;  Indels   31;  Gaps    3;

Qy     798 rgcnttywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaa 857
           :|   |::  ||:| ::| ||:|| ||:||:||::: :  ||:||::| ||:|| ||: |
Db    5187 GGAACTTATATTTAAAAGGGAAGCAGAGCATAAAAGTTTGGAAAATTTGCACCCTGATCA 5246

Qy     858 ygtnathaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtg 917
           :||  |: |:||:||:||:|| ||::: | :|   |:||:||::  || || ||:||:||
Db    5247 TGTGGTAGAAAAGAAAAACCCATTTTCTGGGGAGGAATTCAAGCCGGCTGCAGAAATTTG 5306

Qy     918 yathtgyaaygargarytnaayg-tnaayccncargayaaygggngaraayathwsntgga 976
           :||: |:|| ||:||::  ||:| | ||:   ||:||:| :|| ||:||:  ::    |
Db    5307 CATAAGTAAAGAGGAGCAGAATGTTTAATAGCCAAGACACTGGGGAAAATGGCTAGAAGG 5366

Qy     977 cntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaart 1036
           | | :||::| :: :: ||:||: :  |:  ::  :  |     :  ||
Db    5367 CATTTCAGAGATCTTCACAGCAGCCCCTCCCATTACAGGACACTACTCCC---------- 5416

Qy    1037 ggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccnt 1096
            || :| : |             :: :   |:|        || :|  |: |
Db    5417 TGTGTTTCAGCCACTCCAGCTCCAGCCATGGCTAAAAGGGCCCCAGATATATC------- 5469

Qy    1097 gygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggcarytnc 1156
                     ::  || |  ||  : ||:|  || ::    ||:|| |||||::| |
Db    5470 ------------TTAAGCTGCTGCTCCAGAGGCTGCAAGCTGTAAGCCTTGGCAGCTTC 5516

Qy    1157 cnwsnggngtngarccngtnggngcnaaraaarwsnmgnathgargtntgggarccnccna 1216
           |    || ||  |:|| |  |  ||  |:| ::: :|  |:||:|  |||||:|| ||
Db    5517 CACATGGTGTTAAGCCTGGGGTTGCGCAGAGAACAAGAGTTGAGGCTTGGGAGCCTCCAC 5576

Qy    1217 thmgnttycaraarathtayggnaaycentggatgccnmgncaraarttygcngtnggng 1276
            ::| ||:||:  :|| ||:|| ||:   ||||| | :| ||:||:| :|| |     |
Db    5577 CTAGATTTCAGGGGATGTATGGAAACGTCTGGATGTCCAGGCAGAAGTCTGCAGGACTTG 5636

Qy    1277 tng 1279
           | |
Db    5637 TTG 5639


RESULT   7
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             3.9%;  Score 67;  DB 1;  Length 7218;
  Best Local Similarity   12.6%;  Pred. No. 5.7e-08;
  Matches   64;  Conservative  203;  Mismatches  241;  Indels    0;  Gaps    0;

Qy     580 athahtttyacnggngtnytntayaaygcncarmgngayytnaargargcnatgggngtn 639
           ||:  :| :|   |  | :  |    : |  |::  | : |  |:| : |  ||    |
Db    1536 ATCGCTTCTAGACGCATCTATTCAGTTTCAAAAAACGGCATGTAGGCATCACTGTAATTA 1477

Qy     640 ttygcntgymgngtnccnggnaaytaytaywsnwsnttygaygtngarytncaycaytgy 699
             :  | :             |  |:  :: :  |  |::  ::|    :  :  :
Db    1476 CCTATCTATGCAAGTAGTTAAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRR 1417

Qy     700 aargtnaayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaar 759
           ::|:  :: : : ::   : ::: ::| :|: :   :  :: ::|::|::|: ::: ::|
Db    1416 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1357

Qy     760 carcarwsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmgngar 819
            :| :|:: : : :|::|:  :   :::        ::|:    ::    : ::: ::|
Db    1356 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1297
```

```
Qy    820 gcngarcayaarwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccn 879
          :  ::| : ::|:: :: ::|::      :    :: :: :  : :::|::|::|::
Db   1296 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1237

Qy    880 ttywsngarggnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaay 939
             :: ::|:: ::|   ::|   :  :  ::|: : : : : : :: ::|::|   ::
Db   1236 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1177

Qy    940 gtnaayccncargayaayggngaraayathwsntggacntgycarmgnwsnwsncarcar 999
          :  ::     :|:: :: :: ::|:: : :::  :::    :  :|:: :: ::  :| :|
Db   1176 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1117

Qy   1000 wsnathaarwsnytngcntggmgnccnmgnmgnaartggttytgyggnacnggnccnggn 1059
          :: : :::|::    :   ::::    :: :: ::| ::    : :: :     |  |
Db   1116 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGCAAGCT 1057

Qy   1060 wsnytntgytgygtncarccnmgngayy 1087
            : :    :||:   ||: | :|  |::
Db   1056 CCCTCGACCTGCAGCCAAGCTCGGAATT 1029
```

```
RESULT   8
US-09-106-194-11
; Sequence 11, Application US/09106194
; Patent No. 6262234
;  GENERAL INFORMATION:
;    APPLICANT:  Holloway, James
;    APPLICANT:  Jelinek, Laura
;    APPLICANT:  Durnam, Diane
;    APPLICANT:  Blumberg, Hal
;    TITLE OF INVENTION:  NOVEL NUCLEAR RECEPTOR POLYPEPTIDE
;    TITLE OF INVENTION:  ZPPAR4
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/106,194
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leith, Debra K
;      REGISTRATION NUMBER:  32,619
;      REFERENCE/DOCKET NUMBER:  96-11
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6674
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2067 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-106-194-11
```

Query Match 3.4%; Score 59; DB 4; Length 2067;
Best Local Similarity 9.3%; Pred. No. 3.1e-06;
Matches 101; Conservative 297; Mismatches 676; Indels 12; Gaps 3;

```
Qy    425 cnmgnccnccngargargtnggnccnccnggngcnccnggnytnccncartayacnggng 484
          :  |  :  |      :   :|     :  |     |    |:  |     |:   :  :
Db     13 YTNGARMGNCCNCCNGGNCARGTNWSNYTNGCNGCNCCNTTYYTNCCNGTNACNGARCAR 72

Qy    485 arathwsngaratgacnaartgyccntgyccngayathgarmgnwsngcnttyacngtna 544
          |::|::      ||   :::             |:: |   ::  ::  | : : :
Db     73 AARTTYACNYTNTGGATHMGNCARGCNCARATGWSNTGYGARAAYGCNCAYGCNYTNCCN 132

Qy    545 arytnwsnggnaarytnccnytnccnttyaarccnathathttyacnggngtnytntaya 604
          ::     :       || ::  | :  :::   :|  |: |:: : :  |      :
Db    133 WSNACNCCNCCNTTYTTYYTNTTYTTYYTNTTYCCNTTYTTYCCNYTNG------TNCCN 186

Qy    605 aygcncarmgngayytnaargargcnatgggngtnttygcntgymgngtnccnggnaayt 664
          :: |  ::     : :  |  |::|            : |   : |              :
Db    187 WSNCCNWSNWSNYTNYTNACNAARCCNCCNCCNACNCCNCCNCCNGCNGCNGCNYTNCCN 246

Qy    665 aytaywsnwsnttygaygtngarytncaycaytgyaargtnaayathtggytnatgmgna 724
          : : :  :  :: | ::    :    |  ::  :  ::     : : |     | :
Db    247 CCNCCNCCNCARWSNAARWSNCARAARAARWSNWSNGARCCNGARGCNGTNGCNTTYTTY 306

Qy    725 arcarathytngcnaayaargargarathwsnaarcarcarwsnathcargargtnacnt 784
          |: |:: :    |     ::  :::: :::  |: |  |:::   :  ::|:
Db    307 AARAARCARGCNCAYTGGMGNGARMGNAARMGNAARACNAARCCNAARCARAAYCCNGGN 366

Qy    785 gggtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnwsngaraayy 844
                  ::| :|::  :|:  :          :|::| : ::      |::
Db    367 ACNMGNCARCCNGARCAYTTYTTYTTYACNYTNCCNGARAAYAARCARACNAAYAARCAR 426

Qy    845 tncayccngayaaygtnathaaraaraaraayccnttywsngarggnaarttyaarytng 904
          :  ::   : :::    : |  :: |:: : :   : :  ::    ::  :|  :
Db    427 WSNWSNAARCARWSNCCNCCNACNWSNAARYTNYTNACNMGNMGNMGNMGNCARACNWSN 486

Oy    905 cngcngarathtgyathtgyaaygargarytnaaygtnaayccncargayaayggngara 964
          |     ::  ::|   :     : |:: : |   : :  :  |  :: :: ::
Db    487 CCNTGYWSNCAYGGNGTNMGNYTNAARGGNTGGTTYWSNMGNCARWSNWSNWSNWSNCCN 546

Qy    965 ayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnc 1024
          |:  :   :            | ::  : :  : : :  :: :: :  :   '  | |
Db    547 ACNACNTTYTTYACNMGNGCNGARMGNGAYTTYTGGGCNYTNMGNGGNWSNGGNTGGGAR 606

Qy   1025 cnmgnmgnaartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgng 1084
          | :           ::  : : | :    :          :          |:    | :  :|
Db    607 CARYTNCAYGAYTAYGCNGARMGNGARWSNGGNCAYACNATGMGNGCNCARATHGARGTN 666

Qy   1085 ayytngtnccntgygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarc 1144
          |::   | :    : | :      | :          ::   : |     :    :|
Db    667 ATHCCNTGYAARATHTGYGGNGAYAARWSNWSNGGNATHCAYTAYG---GNGTNATHACN 723

Oy   1145 cntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtnt 1204
            :|   : | :  ::     |  :: :  :    :| :|: |:::     |: :: | :
Db    724 TGYGARGGNTGYAARGGNTTYTTYMGNMGNWSNCARCARAAYAAYGCNWSNTAYWSNTGY 783

Qy   1205 gggarccnccnathmgnttycaraarathtayggnaayccntggatgccnmgncaraart 1264
             :: |  :   ::  ::: | : :    |:    :    : | ||:| :  : : ::
Db    784 CCNMGNCARMGNAAYTGYYTNATHGAYGGNACNAAYMGNAAYMGNTGYCARCAYTGYMGN 843

Oy   1265 tygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaayg 1324
          :: | :  :   :  | :  :          :  :      |    :| : : ::    |:|
Db    844 YTNCARAARTGYYTNGCNYTNGGNATGWSNMGNGAYGCNGTNAARTTYGGNMG---NATG 900

Qy   1325 tnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnm 1384
          :      : : :    :::     | :   : :|  |    :: :: :  ::  :   :
Db    901 WSNAARAARCARMGNGAYWSNYTNTAYGCNGARGTNCARAARCAYCARCARMGNYTNCAR 960

Qy   1385 gnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnc 1444
          | ::  :  | :: :: :       ::   | :|  :     :  :  | ::: :: :
Db    961 GARCARMGNCARCARCARWSNGGNGARGCNGARGCNYTNGCNMGNGTNTAYWSNWSNWSN 1020

Qy   1445 cngaraarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnc 1504
            :: |:::  :  ::   :     :|::  :   ::                |:::    :
Db   1021 ATHWSNAAYGGNYTNWSNAAYYTNAAYAAYGARACNWSNGGNACNTAYGCNAAYGGNCAY 1080

Qy   1505 cntaya 1510
```

```
                |::
Db   1081 GTNATH 1086


RESULT   9
US-08-463-911-6
; Sequence 6, Application US/08463911
; Patent No. 5869330
;  GENERAL INFORMATION:
;    APPLICANT:  Scherer, Philipp E.
;    APPLICANT:  Lodish, Harvey F.
;    TITLE OF INVENTION:  A NOVEL SERUM PROTEIN PRODUCED
;    TITLE OF INVENTION:  EXCLUSIVELY IN ADIPOCYTES
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,911
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  WHI95-05
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1313 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  73..804
US-08-463-911-6


  Query Match             3.0%;  Score 51.8;  DB 2;  Length 1313;
  Best Local Similarity   34.2%;  Pred. No. 0.00022;
  Matches   75;  Conservative   36;  Mismatches  108;  Indels    0;  Gaps    0;

Qy     512 gyccngayathgarmgnwsngcnttyacngtnaarytnwsnggnaarytnccnytnccnt 571
           |: |  |: |: | :| :: || ||:|  ||   ::|        |  |  |  | ||
Db     392 GTGCCTATGTATACCGCTCAGCATTCAGTGTGGGATTGGAGACTTACGTTACTATCCCCA 451

Qy     572 tyaarccnathathttyacnggngtnytntayaaygcncarmgngayytnaargargcna 631
           :| :|| ||:  :||:||    | :| ||:||:   ||::   |::   | |   | |
Db     452 ACATGCCCATTCGCTTTACCAAGATCTTCTACAATCAGCAAAACCACTATGATGGCTCCA 511

Qy     632 tgggngtnttygcntgymgngtnccnggnaaytaytaywsnwsnttygaygtngarytnc 691
             ||    ||:   ||::   | || ||    ||:||::   : | : |: |   : |
Db     512 CTGGTAAATTCCACTGCAACATTCCTGGGCTGTACTACTTTGCCTACCACATCACAGTCT 571

Qy     692 aycaytgyaargtnaayathtggytnatgmgnaarcara 730
           |:       | || ||  |: | :|  | :  ||: | |
Db     572 ATATGAAGGATGTGAAGGTCAGCCTCTTCAAGAAGGACA 610


RESULT  10
US-09-140-804-9
; Sequence 9, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT: Humes, Jacqueline M.
;  TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
;  FILE REFERENCE: 97-49
;  CURRENT APPLICATION NUMBER: US/09/140,804
;  CURRENT FILING DATE: 1998-08-26
;  EARLIER APPLICATION NUMBER: 60/056,983
;  EARLIER FILING DATE: 1997-08-26
;  NUMBER OF SEQ ID NOS: 47
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9
;   LENGTH: 4517
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-140-804-9


  Query Match             3.0%;  Score 51.8;  DB 4;  Length 4517;
  Best Local Similarity   34.2%;  Pred. No. 0.00081;
  Matches   75;  Conservative   36;  Mismatches  108;  Indels    0;  Gaps    0;

Qy     512 gyccngayathgarmgnwsngcnttyacngtnaarytnwsnggnaarytnccnytnccnt 571
           |: |  |: |: | :| :: || ||:|  ||   ::|        |  |  |  | ||
Db     346 gtgcctatgtataccgctcagcattcagtgtgggattggagacttacgttactatcccca 405

Qy     572 tyaarccnathathttyacnggngtnytntayaaygcncarmgngayytnaargargcna 631
           :| :|| ||:  :||:||    | :| ||:||:   ||::   |::   | |   | |
Db     406 acatgcccattcgctttaccaagatcttctacaatcagcaaaaccactatgatggctcca 465

Qy     632 tgggngtnttygcntgymgngtnccnggnaaytaytaywsnwsnttygaygtngarytnc 691
             ||    ||:   ||::   | || ||    ||:||::   : | : |: |   : |
Db     466 ctggtaaattccactgcaacattcctgggctgtactactttgcctaccacatcacagtct 525

Qy     692 aycaytgyaargtnaayathtggytnatgmgnaarcara 730
           |:       | || ||  |: | :|  | :  ||: | |
Db     526 atatgaaggatgtgaaggtcagcctcttcaagaaggaca 564


RESULT  11
US-09-053-866-3/c
; Sequence 3, Application US/09053866
; Patent No. 6111075
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Wenfeng
;    APPLICANT:  Presnell, Scott R.
;    APPLICANT:  Yee, David P.
;    APPLICANT:  Foster, Donald C.
;    TITLE OF INVENTION:  PROTEASE-ACTIVATED RECEPTOR
;    TITLE OF INVENTION:  PAR4 (ZCHEMR2)
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/053,866
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
```

```
;      NAME:  Leith, Debra K
;      REGISTRATION NUMBER:  32,619
;      REFERENCE/DOCKET NUMBER:  98-10
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6674
;      TELEFAX:  206-442-6678
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1155 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-09-053-866-3


  Query Match             2.9%;  Score 50.6;  DB 3;  Length 1155;
  Best Local Similarity   6.3%;  Pred. No. 0.00043;
  Matches   64;  Conservative  256;  Mismatches  703;  Indels    0;  Gaps    0;

Qy     656 cnggnaaytaytaywsnwsnttygaygtngarytncaycaytgyaargtnaayathtggy 715
           : |: |    :      :   :: ::   ::: : |  | ::   :    |::|::|:
Db    1071 YTGRAANARNCCNGCNCKNACYTTRTCNCKRAAYTCNGCNSWNACRTARTARTADATRAA 1012

Qy     716 tnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcarwsnathcarg 775
             :   :   :: ::  :    :: :: :  :: :|    :|:   :     :  ::
Db    1011 NGGRTCNACRCANSWRTTNARNGTNSWNARNGCNARNSWNGGNACRTANGCNCCRTANAR 952

Qy     776 argtnacntgggtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnw 835
           :     |   |   :     : :|    :|:   :  ::  : ::  : : ::  :|
Db     951 RTTNCCCCANGCNSWNGGNSWNGGRTCNSWRTARTGNARNARNARNARNARRTTNSWNGG 892

Qy     836 sngaraayytncayccngayaaygtnathaaraaraaraayccnttywsngarggnaart 895
             ::|:    |  |    ::     : :  :  :  :      :  :   : |   :
Db     891 NACRAARAANGCNACNGCNSWNGCNARNACNACNGCNGTNARNCKNARNGCRTGNCCRTA 832

Qy     896 tyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaayccncargaya 955
           :  :          :  :     :   :    :  :|::  :   |   :   :
Db     831 NCKNCKNCCNSWNGCNGCNARNGTRTGNARNGTNGCNCCRTARCANARNARCATNGCNAR 772

Qy     956 ayggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytng 1015
           :  | ::: | ::     :     | ::: :  ::   :| : :   :|:    :
Db     771 NARNGGNARRAARCANCCNARNARNGCNARRCANGTRAANGCNGGYTGCCARTGNSWNGC 712

Qy    1016 cntggmgnccnmgnmgnaartggttytgyggnacnggnccnggnwsnytntgytgygtnc 1075
           :   |:: |  :     : : :  : ::   :          |  :       ::
Db     711 YTGNGCRTCNARNGGNARNGCRTCRTGRCANARNACNCKRTCNSWNCKNGCNARNCKRAA 652

Qy    1076 arccnmgngayytngtnccntgygtnccngtnaaywsngcngtngcnwsngarggngcnw 1135
           : :    : :   :        |       :                ::   :   | :
Db     651 NGTYTGNCKYTGNARNGTNARNGGNARNGCNARNGCNGCNGCCATNARCCANGCNGCCAT 592

Qy    1136 snccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgna 1195
           :   :  |    || :      :     :  : :|  :      :  : : :    :  :
Db     591 RCANARNCCNARNGCNARNCKNCKNCCNCKNARNGCNCKNGCNCKNARNGGRTGNACNAR 532

Qy    1196 thgargtntgggarccnccnathmgnttycaraarathtayggnaaycccntggatgccnm 1255
             ::     :: |     : ::     | :: :: :      : :|              :
Db     531 NGCNARRTANCKRTCNARNSWNACNGCNGCNARNARNARNACNSWNCCRTACATRTGNCC 472

Qy    1256 gncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncara 1315
           :   :: :            :  :: :        :     :  || :  :: |  |
Db     471 RTANARNGCNGCNGTNGCNARNCKRCANGCNGCYTCNCCRAANGGCCANCKYTGNCCNCK 412

Qy    1316 arggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgng 1375
            :::  ::    : |  : :          :  :  |  :  :  :  |     :  :  :
Db     411 NARRTGRTANGCDATNCKNGGNGGNARNGCNARNGCNARNARNARRTCNGCNGTNGCNAR 352

Qy    1376 cngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytnc 1435
                 :  ::      |     :  :    :   |        :    ::     |
Db     351 RTTCATNARNARCATNGTNSWNGGNARNCKNGGNGCYTGNGTNGCNARNACCCANARNGC 292

Qy    1436 arcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggnatgg 1495
```

```
           :  :        :  ::             :      |     : :   |     :  :
Db     291 NARNCCRTTNGCNGGNARNCCNACNACNARNACNARNCCRTANARNGCNGGNACNARNCK 232

Qy    1496 arwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtayytny 1555
           ::  |  ||   :  ::  :  :    : ::  :    :   ::    :  |:::  :
Db     231 NGTNGGNACCCANCCNARNARNARNGCNCKNSWNSWRTCNGGNARYTCNARNGTRTCNSW 172

Qy    1556 tncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnm 1615
             |::|  |::   :: | :    :   :   :    ::  |   : | :   |  :  :
Db     171 RTCRTTNGCRCANACYTGNCCNGGRTANCCNCKNGGNGCNGGNARDATNSWNGGNGTNSW 112

Qy    1616 gnttygaytgyccnacnggnttymgnacntayatgggnccngtnccnytntgyttyggnc 1675
           :   :  :  :|       : ::  ::  ::      |        ::    | :   |
Db     111 RTCRTCNCCNCCNCCNGTNSWNCCNSWYTCRTCRTANACNSWNGGNGTYTGNGTNCCNCC 52

Qy    1676 art 1678
           ::
Db      51 NSW 49


RESULT  12
US-09-440-325A-2
; Sequence 2, Application US/09440325A
; Patent No. 6280994
; GENERAL INFORMATION:
;  APPLICANT:  Sheppard, Paul O.
;  TITLE OF INVENTION:  Zacl: A Human Metalloenzyme
;  FILE REFERENCE:  98-79
;  CURRENT APPLICATION NUMBER: US/09/440,325A
;  CURRENT FILING DATE:  1999-11-15
;  NUMBER OF SEQ ID NOS:  3
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 2082
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: This degenerate sequence encodes the amino acid
;   OTHER INFORMATION: sequence of SEQ ID NO:1.
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(2082)
;   OTHER INFORMATION: n = A,T,C or G
US-09-440-325A-2


  Query Match             2.8%;  Score 48.8;  DB 4;  Length 2082;
  Best Local Similarity   10.4%;  Pred. No. 0.0026;
  Matches   82;  Conservative  191;  Mismatches  512;  Indels    0;  Gaps    0;

Qy     710 thtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcarwsna 769
           |::||      :|   ::   |:|  :   |:  ::   |:: :::   ::   |:::
Db     316 ttyggnacncargcncayytnttyaargtnacncarttyaargayccngaygtnaayggn 375

Qy     770 thcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngarcaya 829
           :    ::   ::    :  |  :   |::    : :  ::  ::|   :::  ::   :
Db     376 atgytnwsnaarytncaraayathgayaargcngcnytnwsnaargaygarytnmgngar 435

Qy     830 arwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccnttywsngarg 889
           ::  :: :    :       :   :|:|   |: :: |           | ::  |:
Db     436 tayaaygarytnytngcnnnnytngaratgacntaywsnatggcncargtntgyytnaay 495

Qy     890 gnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaayccnc 949
           | : :  :  :|:     :   :::: | ::: | : : | |      ::      :
Db     496 garggnccntgyytnwsnytngarwsngarytngargtnatggcnacnwsnmgngayaar 555

Qy     950 argayaayggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnathaarw 1009
           ::  : :     :  : |  :  ||        ::|  :  :  ::  ::  :|  |
Db     556 gargarytnytntgggcntggcarggntggcargaygcngtnggnmgncarathtgyacn 615

Qy    1010 snytngcntggmgnccnmgnmgnaartggttytgyggnacnggnccnggnwsnytntgyt 1069
           |      :  :   | :   ::   :  |: |  | :      |  :   |
Db     616 acnttygarcaytaygtngarytnwsnaayaargcngcncarytnaayggngtnnnnaar 675
```

```
Qy   1070 gygtncarccnmgngayytngtnccntgygtnccngtnaaywsngcngtngcnwsngarg 1129
Db    676 gayatgggngcnytntggcaywsnaartaygarwsngayacnytngarcargayytngar 735

Qy   1130 gngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarw 1189
Db    736 mgnytnttycargarytnmgnccnytntayytnaayccncayacntaygtnmgnmgngcn 795

Qy   1190 snmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccntgga 1249
Db    796 ytncaymgncaytaygggnccngarytnathgayytnmgnggnccnathccngcncayytn 855

Qy   1250 tgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtng 1309
Db    856 ytnggngaraayacnytngcncarwsntgggtnaayathytngayccngtnytnccntty 915

Qy   1310 tncaraarggnaaygtnggntgggarccncccncaymgngtnccnwsnggngcnccnwsnw 1369
Db    916 ytnaaraarathccngargaygtnacnaarathatgaargtncarcaytggaarccngar 975

Qy   1370 snmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngayw 1429
Db    976 aarytnatgytngargargcngaracnttyttyacntayytnggnytngcnytnccnccn 1035

Qy   1430 snytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcng 1489
Db   1036 gcnccnccnwsnttytggaaraarytnatgytnatgmgnccnacngayggnmgngargtn 1095

Qy   1490 gnatg 1494
Db   1096 gartg 1100


RESULT 13
US-09-351-414-3
; Sequence 3, Application US/09351414
; Patent No. 6265199
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT: Baindur, Nand
;  APPLICANT: Deisher, Theresa A.
;  APPLICANT: Bishop, Paul D.
;  TITLE OF INVENTION: DISINTEGRIN HOMOLOG
;  FILE REFERENCE: 98-29
;  CURRENT APPLICATION NUMBER: US/09/351,414
;  CURRENT FILING DATE: 1999-07-09
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 2088
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: zdintl amino acid degenerate sequence
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (1)...(2088)
;   OTHER INFORMATION: n is any nucleotide
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(2088)
;   OTHER INFORMATION: n = A,T,C or G
US-09-351-414-3


  Query Match             2.8%;  Score 48.8;  DB 4;  Length 2088;
  Best Local Similarity   11.8%;  Pred. No. 0.0026;
  Matches  140;  Conservative  286;  Mismatches  749;  Indels    8;  Gaps    3;

Qy    544 aarytnwsnggnaarytnccnytnccnttyaarccnathathtttyacnggngtnytntay 603
Db    731 aratgytncaygarttywsnaartaymgncarmgnathaarcarcaygcngaygcngtnc 790

Qy    604 aaygcncarmgngayytnaargargcnatgggngtnttygcntgymgngtnccnggnaay 663
Db    791 ayytnathwsnmgngtnacnttycaytayaarmgnwsnwsnytnwsntayttygarggng 850

Qy    664 taytaywsnwsnttygaygtngarytncaycaytgyaargtnaayathtggytnatgmgn 723
Db    851 tntgywsnmgnacnmgnggngtnggngtnaaygartayggnytnccnatggcngtngcnc 910

Qy    724 aarcarathytngcnaayaargargarathwsnaarcarcarwsnathcargargtnacn 783
Db    911 argtnytnwsncarwsnytngcncaraayytnggnathcartgggarccnwsnwsnmgna 970

Qy    784 tgggtnytnytnaargcnttywsnttyathmgngargcngarcayaarwsnwsngaraay 843
Db    971 arccnaartgygaytgyacngarwsntggggnggntgyathatggargaracnggngtnw 1030

Qy    844 ytncayccngayaaygtnathaaraaraaraayccnttywsngarggnaarttyaarytn 903
Db   1031 sncaywsnmgnaarttywsnaartgywsnathytngartaymgngayttyytncarmgng 1090

Qy    904 gcngcngarathtgyathtgyaaygargarytnaaygtnaayccncargayaayggngar 963
Db   1091 gnggnggngcntgyytnttyaaymgnccnacnaarytnttygarccnacngartgyggna 1150

Qy    964 aayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgn 1023
Db   1151 ayggntaygtnga--rgcnggngargartgygaytgyggnttycaygtngartgytaygg 1208

Qy   1024 ccnmgnmgnaartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgn 1083
Db   1209 nytntgytgyaaraartgywsnytnwsnaaygggngcncaytgywsng-ayggnccntgyt 1267

Qy   1084 gayytngtnccntgygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaar 1143
Db   1268 gyaayaayacnwsntgyytnttycarccnmgnggntaygartgymgngaygcngtnaayg 1327

Qy   1144 ccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtn 1203
Db   1328 artgygayathacngartaytgyacnggngaywsnggncartgyccnccnaayytncaya 1387

Qy   1204 tgggarccnccnathmgnttycaraarathtayggnaaycentggatgccnmgncaraar 1263
Db   1388 arcargayggntaygcntgyaaycaraaycarggnmgntgytayaayggngartgyaara 1447

Qy   1264 ttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaay 1323
Db   1448 cnmgngayaaycartgycartayathtggggnacnaargcngcnggnwsngayaarttyt 1507

Qy   1324 gtnggngtgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgn 1383
Db   1508 gytaygaraarytnaayacngarggnacngaraarggna-----aytgyggnaargaygg 1562

Qy   1384 mgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtn 1443
Db   1563 ngaymgntggathcartgywsnaarcaygaygtnttytgyggnttyytnytntgyacnaa 1622

Qy   1444 ccngaraarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtn 1503
Db   1623 yytnacnmgngcnccnmgnathggncarytncarggngarathathccnacnwsnttyta 1682

Qy   1504 ccntayaaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgy 1563
Db   1683 ycaycarggnmgngtnathgaytgywsnggngcncaygtngtnytngaygaygayacnga 1742

Qy   1564 caygayytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygay 1623
Db   1743 ygtnggntaygtngargayggnacnccntgyggnccnwsnatgatgtgyytngaymgnaa 1802

Qy   1624 tgyccnacnggnttymgnacntayatgggnccngtnccnytntgyttyggncarttytty 1683
Db   1803 rtgyytncarathcargcnytnaayatgwsnwsntgyccnytngaywsnaarggnaargt 1862

Qy   1684 ccnttyggnacngcngtnttyacncartgyytntayytncayt 1726
```

```
           ::: |  |        |     |::|     |: | :|::
Db   1863 ntgywsnggncayggngtntgywsnaaygargcnacntgyath 1905


RESULT  14
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17


  Query Match             2.8%;  Score 48.6;  DB 4;  Length 289;
  Best Local Similarity   10.1%;  Pred. No. 0.00038;
  Matches   27;  Conservative  101;  Mismatches  139;  Indels    0;  Gaps    0;

Qy    743 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnt 802
          ::: | ::::::: :|||:: ||:: : :|:::||   :   :::         :::
Db     19 rarcrurarururururarcrararururarcrararurgrnrnrsrnrnrsrnrnrsrnrn 78

Qy    803 tywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtna 862
           ::    :: : : : |   : | ::: | : :   :::           :   :::
Db     79 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 138

Qy    863 thaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyatht 922
          :: | ::: | ::     ::   :::   ::    ::      :   ::: : ::: :
Db    139 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 198

Qy    923 gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 982
          :::   ::: |   :     :       | : :  |: :|||:::: :   :    | |
Db    199 rsrnrnrsrnrnrsrnrnrsrcrargrcrurgrcrgrurararcrurcrururgrcrgrc 258

Qy    983 armgnwsnwsncarcarwsnathaarw 1009
          : :  :  :   |: |::  | :||::
Db    259 ruaaaaaaaaaaaaaaaaaaaaaaaa 285


RESULT  15
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350007
;  CURRENT APPLICATION NUMBER: US/09/244,796
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  EARLIER APPLICATION NUMBER: 09/007,005
;  EARLIER FILING DATE: 1998-01-14
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;   LENGTH: 289
;   TYPE: RNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Translation template
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(289)
;   OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17


  Query Match             2.8%;  Score 48.6;  DB 4;  Length 289;
  Best Local Similarity   10.1%;  Pred. No. 0.00038;
  Matches   27;  Conservative  101;  Mismatches  139;  Indels    0;  Gaps    0;

Qy    743 argargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcnt 802
          ::: | ::::::: :|||:: ||:: : :|:::||   :   :::         :::
Db     19 rarcrurarururururarcrararururarcrararurgrnrnrsrnrnrsrnrnrsrnrn 78

Qy    803 tywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtna 862
           ::    :: : : : |   : | ::: | : :   :::           :   :::
Db     79 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 138

Qy    863 thaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyatht 922
          :: | ::: | ::     ::   :::   ::    ::      :   ::: : ::: :
Db    139 rsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrnrsrnrn 198

Qy    923 gyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyc 982
          :::   ::: |   :     :       | : :  |: :|||:::: :   :    | |
Db    199 rsrnrnrsrnrnrsrnrnrsrcrargrcrurgrcrgrurararcrurcrururgrcrgrc 258

Qy    983 armgnwsnwsncarcarwsnathaarw 1009
          : :  :  :   |: |::  | :||::
Db    259 ruaaaaaaaaaaaaaaaaaaaaaaaa 285


Search completed: July  8, 2002, 16:22:51
Job time: 12295 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 8, 2002, 18:50:56 ; Search time 3393.27 Seconds
(without alignments)
6885.167 Million cell updates/sec

Title: US-09-997-610-7
Perfect score: 1731
Sequence: 1 atgytnathgargcntaygg.....gyytntayytncaytgyatg 1731

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	479.2	27.7	2615	10	BE420422	BE420422 32-393 hu
	2	450.2	26.0	2009	11	BC004496	BC004496 Homo sapi
	3	398	23.0	1514	10	BE512633	BE512633 32-1514 h
	4	336.8	19.5	870	10	BI488505	BI488505 603021014
	5	335.6	19.4	1050	10	BM472108	BM472108 AGENCOURT
	6	321.8	18.6	1036	10	BM471183	BM471183 AGENCOURT
	7	318.6	18.4	898	10	BM457166	BM457166 AGENCOURT
	8	313	18.1	717	12	AG086951	AG086951 Pan trogl
c	9	306.6	17.7	736	10	BI754555	BI754555 603023384
	10	300.4	17.4	654	12	AG112666	AG112666 Pan trogl
	11	296	17.1	676	12	AG064424	AG064424 Pan trogl
	12	287.4	16.6	691	12	AG062447	AG062447 Pan trogl
	13	286.2	16.5	720	12	AG116938	AG116938 Pan trogl
	14	283.8	16.4	623	12	AQ037711	AQ037711 CIT-HSP-2
	15	283.2	16.4	677	12	AG097258	AG097258 Pan trogl
	16	282.4	16.3	704	12	AQ588828	AQ588828 CITBI-E1-
	17	281.6	16.3	659	12	AG092858	AG092858 Pan trogl
	18	281.4	16.3	701	12	AQ588873	AQ588873 CITBI-E1-
	19	281	16.2	729	10	BM045873	BM045873 603624875
	20	279.6	16.2	760	10	BG281182	BG281182 602402107
	21	279.4	16.1	697	12	AG018926	AG018926 Homo sapi
	22	278.8	16.1	885	10	BG761979	BG761979 602718930
	23	277.2	16.0	1182	10	BM451317	BM451317 AGENCOURT
	24	276.2	16.0	639	12	AG115932	AG115932 Pan trogl
	25	275.2	15.9	676	12	AG125093	AG125093 Pan trogl
	26	275	15.9	721	10	BE728616	BE728616 601563827
	27	274.8	15.9	671	12	AG045131	AG045131 Pan trogl
	28	274.6	15.9	877	10	BI916691	BI916691 603178655
	29	272.2	15.7	625	12	AG070563	AG070563 Pan trogl
	30	271.4	15.7	704	12	AQ020199	AQ020199 CIT-HSP-2
	31	271.2	15.7	686	12	AG119684	AG119684 Pan trogl
	32	268.8	15.5	660	12	AG057341	AG057341 Pan trogl
c	33	268.6	15.5	746	12	AG001408	AG001408 Homo sapi
	34	266	15.4	623	12	AG075663	AG075663 Pan trogl
c	35	266	15.4	699	9	AL602279	AL602279 DKFZp686B
	36	264.8	15.3	543	12	B94403	B94403 CIT-HSP-217
	37	263.8	15.2	675	12	AG096655	AG096655 Pan trogl
	38	261.4	15.1	833	9	AU140358	AU140358 AU140358
	39	261	15.1	771	12	AG001229	AG001229 Homo sapi
	40	260.6	15.1	716	12	AG104336	AG104336 Pan trogl
c	41	260.6	15.1	940	10	BG679573	BG679573 602627616
	42	260.6	15.1	955	10	BG681769	BG681769 602627877
	43	257.4	14.9	649	9	AW857140	AW857140 RC1-CT030
	44	256.6	14.8	900	10	BM449699	BM449699 AGENCOURT
	45	255.4	14.8	676	12	AQ239408	AQ239408 CIT-HSP-2

ALIGNMENTS

RESULT 1
BE420422
LOCUS BE420422 2615 bp mRNA linear EST 11-OCT-2001
DEFINITION 32-393 human bone marrow cDNA library Homo sapiens cDNA, mRNA sequence.
ACCESSION BE420422
VERSION BE420422.1 GI:16041640
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2615)
AUTHORS Lu,X., Cui,L. and Li,Y.
TITLE DDRT-PCR from B cell
JOURNAL Unpublished (2000)
COMMENT Contact: xingwu Lu,liangxian Cui,yonghai Li
Department of Biochemistry
Institute of Basic Medical Science, Peking Union Medical College
DongDan SanTiao 5, Beijing, P.R.C, 100005
Tel: 86-010-65296951
Email: luxingwu@263.net
full-length and coding sequence.
FEATURES Location/Qualifiers
source 1. .2615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="human bone marrow cDNA library"
/tissue_type="bone marrow"
BASE COUNT 705 a 531 c 641 g 738 t
ORIGIN

Query Match 27.7%; Score 479.2; DB 10; Length 2615;
Best Local Similarity 47.1%; Pred. No. 2.6e-99;
Matches 500; Conservative 222; Mismatches 299; Indels 41; Gaps 4;

```
Qy    706 aayathtggytnatgmgnaarcarathytngcnaayaargargarathwsnaarcarcar 765
          ||: |:  |:|   |:|  | ||  |:          :||:||:||::: ||:||:||:
Db    494 AACCTTGAGCTTGAGAGAGATCATTTAGGGTATCTGGCAGAAGAAATTTCTAAGCAGCAA 553
```

```
Qy    766 wsnathcargargtnacntgggtnytnytnaargcnttywsnttyathmgngargcngar 825
          :: ||:||:||:|| || ||||| :| :| ||:|| ||::: ||:||::| ||:|| | :
Db    554 AGCATTCAAGAGGTAACTTGGGTGCTGTTAAAGGCATTTAGTTTTATAAGGGAAGCAGGG 613

Qy    826 cayaarwsnwsngaraayytncayccngayaaygtnathaaraaraaraayccnttywsn 885
          ||: |::: :  ||:||::| || || ||:||:|  ||: |:||:||:| :|| | :::
Db    614 CATGAAAGTTTGGAAAATTTGCAGCCTGACAATGCAATAGAAAAGAAAATCCCATATTCT 673

Qy    886 garggnaarttyaarytngcngcngarathtgyathtgyaaygargarytnaaygtnaay 945
          ||:|  ||: |:||::| || || ||:||:||:||:||:||: |:||::| ||:|| ||:
Db    674 GAGGAGAAAATCAAGCTGGCTGCAGAAATTTGCATATGTAACAAGGAGTTAAATGTTAAT 733

Qy    946 ccncargayaayggngaraayathwsntggacntgycarmgnwsnwsncarcarwsnath 1005
          || ||:||:||:|| ||:||: |:::  || | ||:||::| :: :: ||:||: :  |:
Db    734 CCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGCCAGAGGTCTTCACAGCAGCCCCTC 793

Qy   1006 aarwsnytngcntggmgnccnmgnmg---naartggttytgyggnacnggnccnggnwsn 1062
            ::: :  || || :| || :| :(    ||:|| | :  :||  | || || || ::
Db    794 CCATCACAGGCCTGCAGGCCTAGGAGGAAAAAGTGATCTCTTGGGCCAGGCCCAGGGTCC 853

Qy   1063 ytntgytgygtncarccnmgngayytngtnccntgygtnccngtnaaywsngc------- 1115
          :  ||:||:|| ||:||  | ||::| || || ||:|| || |    : : ||
Db    854 CCATGCTGTGTGCAGCCTGGGGACTTGGTGCCTTGTGTCCCAGCTGCTCCAGCCATGGCT 913

Qy   1116 -------------------------ngtngcnwsngarggngcnwsnccnaarccntgg 1149
                                     || || :: ||:|| || ::    ||:|| |||
Db    914 GAAAGGGGCCAACATAGAGCTCGGATTGTGGCTTCAGAGGGTGCAAGCTGTAAGCCTTGG 973

Qy   1150 carytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggar 1209
          ||::| || :: || || ||:|| |   | ||  |:||::: :  ||:| :|| |||||:
Db    974 CAGCTTCCATGTGGTGTTGAGCCTGCCAGTGCACAGAAGTCAAAAATTGGGGTTTGGGAA 1033

Oy   1210 ccnccnathmgnttycaraarathtayggnaaycentggatgccnmgncaraarttygcn 1269
          || |    ::| ||:||:||:|| ||:|| ||: | |||||||  | ||:||:| :||
Db   1034 CCTCTGCCTAGATTTCAGAAGATGTATGGAAATGCCTGGATGCCCTGGCAGAAGTCTGCT 1093

Qy   1270 gtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggn 1329
          || || |  ||  : :: |||:| |  :: || :| |  || ||:||:|| ||:|| ||
Db   1094 GTAGGGGAGGGGCCCTCATGGAGAATCTCTGCTAGGGCAGTACAGAAGGGAAATGTGGGA 1153

Qy   1330 tgggarccncccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsn 1389
          | |||:|| |  ||::|  | || :: || || ||  : ::  | || || :| :|  :
Db   1154 TCGGAGCCCCAGCACAGAATCCCTACTGGGGCACCGCCTAGTGGAGCTGTGAGAAGAGGG 1213

Qy   1390 ccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngar 1449
          || || :: :: :| :  ||:|| |  :| :: || ||::: :| |  ||:|  || | :
Db   1214 CCACCATCCTCCAGACCCCAGAATGATAGATCCACTGACAGCTTGCTCCATGCACCTGGA 1273

Qy   1450 aarwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntay 1509
          ||:::  | ||    ||:  :||:||  | ||:|| ||  |   |||: : || || | :
Db   1274 AAAACCGCAGA----CAACACCAGCCCATGAAAGCAGCTAGGCAGGAGGCTGTACCCTGC 1329

Qy   1510 aaracngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygay 1569
          ||: |    |  || ||::|  | ||:|   | || | : |::| :| ||:   ||:||:
Db   1330 AAAGCCACAGGGGCGGAGCTGCCCAAGATCATGGGAACCCACCTCTTGCATCAGCATGAC 1389

Qy   1570 ytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccn 1629
          :| ||:|| :|   :|| || ||: | ||:||:||:|| || :| :| ||:||:||:||
Db   1390 CTGGATGTGAGATTTGGAGTTAAAGGAGATCATTTTGGAGCTTTAAGATTTGACTGCCCC 1449

Qy   1630 acnggnttymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccntt 1688
           |  | ||::  || | :|||||| || ||  || :| ||:||:   ||:||: |:|| ||
Db   1450 TCTAGATTTCAGACTTGCATGGGGCCTGTAGCCCCTTTGTTTTACTCAATTTCTCCCATT 1509

Oy   1689 yggnacngcngtnttyacncartgyytntayytncaytgyat 1730
          :|| || || || ||:|| ||:| ::| ||::  ||:||:||
Db   1510 TGGAACAGCTGTGTTTACACAATACCTGTACCCCCATTGTAT 1551


RESULT   2
BC004496
LOCUS       BC004496               2009 bp    mRNA    linear   HTC 12-JUL-2001
DEFINITION  Homo sapiens, Similar to hypothetical protein FLJ14058, clone
            IMAGE:3831313, mRNA.
ACCESSION   BC004496
VERSION     BC004496.1  GI:14709139
KEYWORDS    HTC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2009)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 14 Row: d Column: 7
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1. .2009
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3831313"
                     /tissue_type="Skin, melanotic melanoma."
                     /clone_lib="NIH_MGC_20"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
BASE COUNT      556 a    429 c    513 g    511 t
ORIGIN


  Query Match             26.0%;  Score 450.2;  DB 11;  Length 2009;
  Best Local Similarity   45.2%;  Pred. No. 1.1e-92;
  Matches  452;  Conservative  214;  Mismatches  321;  Indels   13;  Gaps    3;

Qy    744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
          :||:||:||::: | :||:||::: ||:||:||: | || |||||  :| :| ||:|| ||
Db    489 AGAAGAAATTTCTATGCAGCAAAGCATTCAAGAGTTGACTTGGGTGCTGTTAAAGGCATT 548

Qy    804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
          ::: ||:||::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||
Db    549 CAGATTTATAAGGGAAGCAGAGCATAAAAGTTCAGAAAATTTGCAGCCTGACAATGTGAT 608

Qy    864 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
          : |:||:||:| :|| ||::  ||:|  ||:||:||::  || ||  |:||:||:||: |
Db    609 AGAGAAGAAAAGCCCATTTTTTGAGGAGAAATTCAAGCCAGCTGCATAAATTTGCATAAG 668

Qy    924 yaaygargarytnaaygtnaaycccncargayaayggngaraayathwsntggacntgyca 983
          :||: |: |::   |:|| ||:|| ||:||:||:|| || |:||:::  |||| | :||
Db    669 TAACAAGAAGCCGCATGTTAATCCCCAAGACAACGGAGAAGACATCTCCAGGACATATCA 728

Qy    984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 1043
          ::| :: :: ||:||: :  |:  ::: :  ||   |:| || :| :| ||:|||||:
Db    729 GAGGTCTTCACAGCAGCCCCTCCCGTCACAGGCCCAGAGGCCTAGGAGAAAATGGTTTCC 788

Qy   1044 yggnacnggnccnggnwsnytnt------gytgygtncarccnmgngayytngtncc--- 1094
```

```
Db    789 TGGACCTGTGCAGCCTAGGTACTTGGCACCCTGTGTCCCAGCTACTCCAGTAGTGGCTGA 848

Qy   1095 ---ntgygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggca 1151
Db    849 AAGGGGCCAAAGTACAGCTTACGCCATGGCTTCAGAGGGTGCAAGCGCCAAGCCTTGGCA 908

Qy   1152 rytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarcc 1211
Db    909 GCTTCCGCATGCTGTTGAGCCTGCAAATGCACAGACGTCAAGAATTTGGATTTGGGAACC 968

Qy   1212 nccnathmgnttycaraarathtayggnaaycentggatgccnmgncaraarttygcngt 1271
Db    969 TCTGCCTAGATTTCAGGAGATGTACGGAAGTGTCCGGATGCCCAGGCAAAAGTTTGCTGC 1028

Qy   1272 nggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntg 1331
Db   1029 AGGGGCAGGGCCTTCATGGAGAACCTCTGCTAGGGCAGTGTGGAAGGGAAACATGGGGTC 1088

Qy   1332 ggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsncc 1391
Db   1089 TGAGCCCCCACACAGAATCCCTACTGAGGCACCACCTAGTGGAGCTGTGAGAAAAGAGCC 1148

Qy   1392 nccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraa 1451
Db   1149 ACCATCCTCCACACCCCAGAATGATAGATCCACCAACAGCTTGCACCATGCACCTGGAAA 1208

Qy   1452 rwsnacngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntayaa 1511
Db   1209 AGCCGCAGACATTCAACACCAGCCCGTGAAAGCAGCCAAAAGGGGGAATATATCCTGCAA 1268

Qy   1512 racngtngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygayyt 1571
Db   1269 AGCCACAGGGGAGGAGCTTCCCAAGACTATGGGAACCTACCTCTTACATCAGCACGACCT 1328

Qy   1572 ngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnac 1631
Db   1329 TTGTGTGAGACCTGGAGTCAAAGGAGATCATTTTGGAGCTTTAAAATCTGACTGCCCTGC 1388

Qy   1632 nggnttymgnacntayatgggncc-ngtnccnytntgyttyggncarttyttyccnttyg 1690
Db   1389 TGGATTTCAGACTTGCATGGGCCCTACAACCCCTTTGTTTTGGCCGATTTCTCCAATTTG 1448

Qy   1691 gnacngcngtnttyacncartgyytntayytncaytgyat 1730
Db   1449 GAAAGGCTGTATTTACCCAATACCTGTACCCCCATTGTAT 1488
```

```
RESULT   3
BE512633
LOCUS       BE512633                1514 bp    mRNA    linear   EST 11-OCT-2001
DEFINITION  32-1514 human B lymphocyte cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   BE512633
VERSION     BE512633.1  GI:16041645
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Lu,X., Cui,L. and Li,Y.
  TITLE     DDRT-PCR from B cell
  JOURNAL   Unpublished (2000)
COMMENT     Contact: xingwu Lu,liangxian Cui,yonghai Li
            Department of Biochemistry
            Institute of Basic Medical Science, Peking Union Medical College
            DongDan SanTiao 5, Beijing, P.R.C, 100005
            Tel: 86-010-65296951
            Email: luxingwu@263.net
            full cDNA sequence.
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="human B lymphocyte cDNA library"
                     /tissue_type="bone marrow"
BASE COUNT      380 a    341 c    398 g    395 t
ORIGIN
```

Query Match 23.0%; Score 398; DB 10; Length 1514;
Best Local Similarity 47.1%; Pred. No. 1.1e-80;
Matches 427; Conservative 182; Mismatches 256; Indels 41; Gaps 4;

```
Qy    862 athaaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyath 921
Db      1 ATAGAAAAGAAAATCCCATATTCTGAGGAGAAAATCAAGCTGGCTGCAGAAATTTGCATA 60

Qy    922 tgyaaygargarytnaaygtnaayccncargayaaygggngaraayathwsntggacntgy 981
Db     61 TGTAACAAGGAGTTAAATGTTAATCCCCAAGACAATGGGGAAAATGTCTCCAGGGCATGC 120

Qy    982 carmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---naartgg 1038
Db    121 CAGAGGTCTTCACAGCAGCCCCTCCCATCACAGGCCTGCAGGCCTAGGAGGAAAAAGTGA 180

Qy   1039 ttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtnccntgy 1098
Db    181 TCTCTTGGGCCAGGCCCAGGGTCCCCATGCTGTGTGCAGCCTGGGGACTTGGTGCCTTGT 240

Qy   1099 gtnccngtnaaywsngc----------------------------------ngtngcnwsn 1125
Db    241 GTCCCAGCTGCTCCAGCCATGGCTGAAAGGGGCCAACATAGAGCTCGGATTGTGGCTTCA 300

Qy   1126 garggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaar 1185
Db    301 GAGGGTGCAAGCTGTAAGCCTTGGCAGCTTCCATGTGGTGTTGAGCCTGCCAGTGCACAG 360

Qy   1186 aarwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggnaayccn 1245
Db    361 AAGTCAAAAATTGGGGTTTGGGAACCTCTGCCTAGATTTCAGAAGATGTATGGAAATGCC 420

Qy   1246 tggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgn 1305
Db    421 TGGATGCCCTGGCAGAAGTCTGCTGTAGGGGAGGGGCCCTCATGGAGAATCTCTGCTAGG 480

Qy   1306 gtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccn 1365
Db    481 GCAGTACAGAAGGGAAATGTGGGATCGGAGCCCCAGCACAGAATCCCTACTGGGGCACCG 540

Qy   1366 wsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacn 1425
Db    541 CCTAGTGGAGCTGTGAGAAGAGGGCCACCATCCTCCAGACCCCAGAATGATAGATCCACT 600

Qy   1426 gaywsnytncarcaygtnccngaraarwsnacngayacncartgycarccngtnaargcn 1485
Db    601 GACAGCTTGCTCCATGCACCTGGAAAAACCGCAGA----CAACACCAGCCCATGAAAGCA 656

Qy   1486 gcnggnatggarwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggn 1545
Db    657 GCTAGGCAGGAGGCTGTACCCTGCAAAGCCACAGGGGCGGAGCTGCCCAAGATCATGGGA 716

Qy   1546 athtayytnytncaytgycaygayytngaygtnmgncayggngtnaarmgngaycaytty 1605
Db    717 ACCCACCTCTTGCATCAGCATGACCTGGATGTGAGATTTGGAGTTAAAGGAGATCATTTT 776

Qy   1606 ggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnyt 1664
Db    777 GGAGCTTTAAGATTTGACTGCCCCTCTAGATTTCAGACTTGCATGGGGCCTGTAGCCCCT 836

Qy   1665 ntgyttyggncarttyttyccnttyggnacngcngtnttyacncartgyytntayytnca 1724
Db    837 TTGTTTTACTCAATTTCTCCCATTTGGAACAGCTGTGTTTACACAATACCTGTACCCCCA 896

Qy   1725 ytgyat 1730
```

```
          :||:||
Db    897 TTGTAT 902


RESULT   4
BI488505
LOCUS       BI488505                870 bp    mRNA    linear   EST 28-AUG-2001
DEFINITION  603021014F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5191803 5',
            mRNA sequence.
ACCESSION   BI488505
VERSION     BI488505.1  GI:15327733
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 870)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11479  row: k  column: 04
            High quality sequence stop: 849.
FEATURES             Location/Qualifiers
     source          1. .870
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5191803"
                     /clone_lib="NIH_MGC_114"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is  destroyed
                     upon cloning).  Average insert size 1.5 kb, insert size
                     range 1-3 kb.  Library is normalized and enriched for
                     full-length  clones and was constructed by C. Gruber
                     (Invitrogen).  Research  Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT      222 a    211 c    256 g    181 t
ORIGIN


  Query Match             19.5%;  Score 336.8;  DB 10;  Length 870;
  Best Local Similarity   44.6%;  Pred. No. 1e-66;
  Matches  378;  Conservative  161;  Mismatches  273;  Indels   35;  Gaps    3;

Qy    847 cayccngayaaygtnathaaraaraaraayccnttywsngarggnaarttyaarytngcn 906
          || || ||: |:|   |: |:||:||:||:|| ||::  ||:|  ||:||:||::  ||
Db      2 CAGCCTGATGATGAAGTAGAAAAGAAAAACCCATTTTTTGAGGATAAATTCAAGCCAGCT 61

Qy    907 gcngarathtgyathtgyaaygargarytnaaygtnaayccncargayaaygggngaraay 966
          || ||:||:||:||:||:| : |:||::  ||:|| ||:|  ||:||: |:|| ||:||:
Db     62 GCAGAAATTTGCATATGTAGCAAGGAGTCTAATGTTAATCTCCAAGACCATGGGGAAAAT 121

Qy    967 athwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnc-c 1025
           |:::  || | ||:||::|  : ::   :||: :  |:  ::: :  ||  |  | |
Db    122 GTCTCCGGGCCATGTCAGAGACCTTCATGGCAGCCCCTCCCATCACAGGCTGGAGGCCAG 181

Qy   1026 nmgnmgnaartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgnga 1085
          :| :  ||: ||||: |:||  | || || |  :: :  ||:| :|| ||:|| :| ||
Db    182 GAGGAAAAAGAGGTTTCGTGGGCCAGGCCCAGCATCCCCGTGCTTTGTGCAGCCTAGGGA 241

Qy   1086 yytngtnccntgygtnccngt----------------------------------naayws 1112
          ::| || || ||:|| || ||                                  : :

Db    242 CTTGGTGCCCTGTGTCCCAGTCGCTCCAGCTGTGACTGAAAGGGGCCAACGTAGAGCTCG 301

Qy   1113 ngcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarcc 1172
          || || || :: ||:|| |  ::  | ||:|  ||||| :| ||  : || || ||:||
Db    302 GGCTGTGGCTTCAGAGGGTGGAAGCTCTAAGCTTTGGCAGCTTCCACGTGGTGTTGAGCC 361

Qy   1173 ngtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarat 1232
           || ||  |:||::: :| | :||:|| ||||| :|| |    ::| ||:||:||:||
Db    362 TACAGGTGCACAGAAGTCAAGAACTGAGGTTTGGGAACCTCTGCCTAGATTTCAGAAGAT 421

Qy   1233 htayggnaayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmg 1292
           | :|| ||: | |||||||| :| ||:||:||:|| |  || || ||  : :  |||:|
Db    422 GTTTGGAAATGCCTGGATGCCCAGGCAAAAGTTTGCTGCAGGAGTGGGGCCCTTATGGAG 481

Qy   1293 nacnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtncc 1352
          || :: ||  | |  || ||:||:|  ||:|| || | |||:|| |  || :| || |
Db    482 AACCTCTGCTGGAGCAGTGCAGAAGGAAAATGTGGGATCGGAGCCCCGACAGAGAGTCCA 541

Qy   1353 nwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraa 1412
           :: || || |   : ::  |  | || :| :|    || || :: :: :| :  ||:||
Db    542 TACTGGGGCACTGCCTAGTGGAACTGTGAGAAGACAGCCACCGTCCTCCAGACCCCAGAA 601

Qy   1413 rggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgyca 1472
          || :| :: ||  |::: :| || ||:|| || | :||:    | | :|| ||:||:||
Db    602 TGGTAGATCCACCAACAGCTTGCACCATGTGCCTGGAAAAGATGCAGGTACTCAATGCCA 661

Qy   1473 rccngtnaargcngcngggnatggarwsngtnccntayaaracngtngtngcngarytnac 1532
          :||  | ||:|| ||  | | |||: : || || | :||: |    |    ||::|  |
Db    662 GCCCATGAAGGCAGCCAGGAGGGAGGCTGTACCCTGCAAAGCCACAGGGAGGGAGCTGTC 721

Qy   1533 naaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaa 1592
          ||:||  | || | : |::  :    : |:     ::| ||:|  :| | :|| || ||
Db    722 CAAGACCATGGGAACCCACTCTTGCATCAGCGTGACCTTGGATGCGAGACCTGGAGTCAA 781

Qy   1593 rmgngaycayttyggngcnyt-nmgnttygaytgyccnacnggnttymgnacntayatgg 1651
          : | ||:||:||:|| || :|  :  | :||:   ||  | || ||::  || | :|| |
Db    782 AGGAGATCATTTTGGAGCTTTAACAATGTGACCTGCCTGCTGGATTTCAGACTTGCATTG 841

Qy   1652 gnccngt 1658
           |  |  |
Db    842 GGGCCCT 848


RESULT   5
BM472108
LOCUS       BM472108                1050 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6465359 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5539381
            5', mRNA sequence.
ACCESSION   BM472108
VERSION     BM472108.1  GI:18521150
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1050)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12233  row: m  column: 14
            High quality sequence stop: 658.
FEATURES             Location/Qualifiers
     source          1. .1050
                     /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5539381"
                     /clone_lib="NIH_MGC_72"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT       295 a    262 c    272 g    221 t
ORIGIN

  Query Match             19.4%;  Score 335.6;  DB 10;  Length 1050;
  Best Local Similarity   43.8%;  Pred. No. 2.2e-66;
  Matches  335;  Conservative  169;  Mismatches  248;  Indels   12;  Gaps    2;

Qy    744 rgargarathwsnaarcarcarwsnathcargargtnacntgggtnytnytnaargcntt 803
          :||:||:||:::.| :||:||::: ||:||:||: | || ||||| :| :| ||:|| ||
Db     16 AGAAGAAATTTCTATGCAGCAAAGCATTCAAGAGTTGACTTGGGTGCTGTTAAAGGCATT 75

Qy    804 ywsnttyathmgngargcngarcayaarwsnwsngaraayytncayccngayaaygtnat 863
          ::: ||:||::| ||:|| ||:||:||::: :: ||:||::| || || ||:||:|| ||
Db     76 CAGATTTATAAGGGAAGCAGAGCATAAAAGTTCAGAAAATTTGCAGCCTGACAATGTGAT 135

Qy    864 haaraaraaraayccnttywsngarggnaarttyaarytngcngcngarathtgyathtg 923
          : |:||:||:| :|| ||::  ||:|  ||:||:||::  || ||  |:||:||:||: |
Db    136 AGAGAAGAAAAGCCCATTTTTTGAGGAGAAATTCAAGCCAGCTGCATAAATTTGCATAAG 195

Qy    924 yaaygargarytnaaygtnaayccncargayaayggngaraayathwsntggacntgyca 983
          :||: |: |::   |:|| ||:|| ||:||:||:|| || |:||:::  |||| | :||
Db    196 TAACAAGAAGCCGCATGTTAATCCCCAAGACAACGGAGAAGACATCTCCAGGACATATCA 255

Qy    984 rmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmgnaartggttytg 1043
          ::| :: :: ||:||: :  |:  ::: :  ||   |:| || :| :| ||:||||||:
Db    256 GAGGTCTTCACAGCAGCCCCTCCCGTCACAGGCCCAGAGGCCTAGGAGAAAATGGTTTCC 315

Qy   1044 yggnacnggnccnggnwsnytnt------gytgygtncarccnmgngayytngtncc--- 1094
          :||  | |  |     :: :  |          :||:|| | : | :          | ||  |
Db    316 TGGACCTGTGCAGCCTAGGTACTTGGCACCCTGTGTCCCAGCTACTCCAGTAGTGGCTGA 375

Qy   1095 ---ntgygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggca 1151
             |:          ::  ||  | || :: ||:|| || ::  | ||:|| |||||
Db    376 AAGGGGCCAAAGTACAGCTTACGCCATGGCTTCAGAGGGTGCAAGCGCCAAGCCTTGGCA 435

Qy   1152 rytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarcc 1211
          ::| ||    |  || ||:|| |      ||  |:| ::: :| ||:  : | ||||:||
Db    436 GCTTCCGCATGCTGTTGAGCCTGCAAATGCACAGACGTCAAGAATTTGGATTTGGGAACC 495

Qy   1212 nccnathmgnttycaraarathtayggnaayccntggatgccnmgncaraarttygcngt 1271
             ::| ||:||: |:|| ||:|| | :    |||||||| :| ||:||:||:|| |
Db    496 TCTGCCTAGATTTCAGGAGATGTACGGAAGTGTCCGGATGCCCAGGCAAAAGTTTGCTGC 555

Qy   1272 nggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntg 1331
          || |  ||  : :: |||:| || :: || :| |  ||    :|||:|| ||: | || |
Db    556 AGGGGCAGGGCCTTCATGGAGAACCTCTGCTAGGGCAGTGTGGAAGGGAAACATGGGGTC 615

Qy   1332 ggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsncc 1391
          ||:|| || ||::|  | || :: |  || ||  :  ::  | || || :| :  |  ||
Db    616 TGAGCCCCCACACAGAATCCCTACTGAGGCACCACCTAGTGGAGCTGTGAGAAAAGAGCC 675

Qy   1392 nccnwsnwsnmgnytncaraarggnmgnwsnacngaaywsnytncarcaygtnccngaraa 1451
             || :: :: :  :  ||:|| |  :| :: ||  |::: :| || ||:|  ||   :||
Db    676 ACCATCCTCCACACCCCAGAATGATAGATCCACCAACAGCTTGCACCATGCACCCTGGAA 735

Qy   1452 rwsnacngayacncartgycarccngtnaargcngcnggnatgg 1495
          : :  | ||:|  ||:  :||:|| |   |:|| ||    | ||
Db    736 AGCCGCAGACATTCAACACCAGCCCGGGGAAGCAGCCAAAAGGG 779


RESULT   6
BM471183
LOCUS       BM471183                1036 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6478314 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5563049
            5', mRNA sequence.
ACCESSION   BM471183
VERSION     BM471183.1  GI:18520225
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1036)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            Tissue Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            cDNA Library Arrayed by: Agencourt Bioscience Corporation
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12293  row: g  column: 18
            High quality sequence stop: 661.
FEATURES             Location/Qualifiers
     source          1. .1036
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5563049"
                     /clone_lib="NIH_MGC_72"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT       254 a    239 c    294 g    247 t      2 others
ORIGIN

  Query Match             18.6%;  Score 321.8;  DB 10;  Length 1036;
  Best Local Similarity   45.0%;  Pred. No. 3.4e-63;
  Matches  390;  Conservative  166;  Mismatches  267;  Indels   43;  Gaps    6;

Qy    907 gcngarathtgyathtgyaaygargaryt--naaygtnaayccncargayaayggngara 964
          || ||:| :| :||: |:| : |:||::|   ||:|| ||:|| ||:||: |:|| ||:|
Db      1 GCAGAAAATTACATAAGTAGCAAGGAACTACTAATGTTAATCCCCAAGACCATGGGGAAA 60

Qy    965 ayathwsntggacntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnc 1024
          |: |:::  |  | ||:||::|  : :: | :||:::  |:  ::: :  ||   |:| |
Db     61 ATGTCTCCAGTTCCTGTCAGAGACCTTCACTGCAGACCCTCCCATCACAGGCCCAGAGGC 120

Qy   1025 cnmgnmg---naartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnm 1081
          | :| :|    ||:|||||:||:||    ||  | || :: :  ||:||:||  |:|| :
Db    121 CCAGGAGGAAAAAGTGGTTCTGTGGGCTGGGTTCAGGGTCCCCATGCTGTGTGAAGCCTA 180

Qy   1082 gngayytngtnccntgygtnccng----------------------------------tna 1108
          | ||::| || || ||:|| || |                                   |
Db    181 GGGACTTGGTGCCCTGTGTCCCAGCTGCTCCAGCCTTGGCTGAAAGGGGCTAATGTATAG 240

Qy   1109 aywsngcngtngcnwsngarggngcnwsnccnaarccntggcar---ytnccnwsnggng 1165
             :   || || ||  :  ||:|| |  :: || ||:|| |||||:   :| || :: || |
Db    241 CTCTGGCTGTGGCTTTAGAGGGTGGAAGCCCCAAGCCTTGGCAGCTTCTTCCATGTGGTG 300

Qy   1166 tngarccngtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttyc 1225
          | ||:|| || || ||  |:||::: :| | :||:|| ||||:|| |   | :| ||:|
Db    301 TTGAGCCTGTGGGTGCACAGAAGTCAAGAACTGAGGTTTGGGAACCTCTGCT-AGATTCC 359

Qy   1226 araarathtayggnaaygccntggatgccnmgncaraarttygcngtnggngtnggnwsnw 1285
          |:||:|| ||:|| ||: | |||||||| :| ||:||:||:|| |   | | || : :
Db    360 AGAAGATGTATGGAAATGCCTGGATGCCCAGGCAAAAGTTTGCTGCAAGGTTGGGGCCCT 419

Qy   1286 sntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaym 1345
```

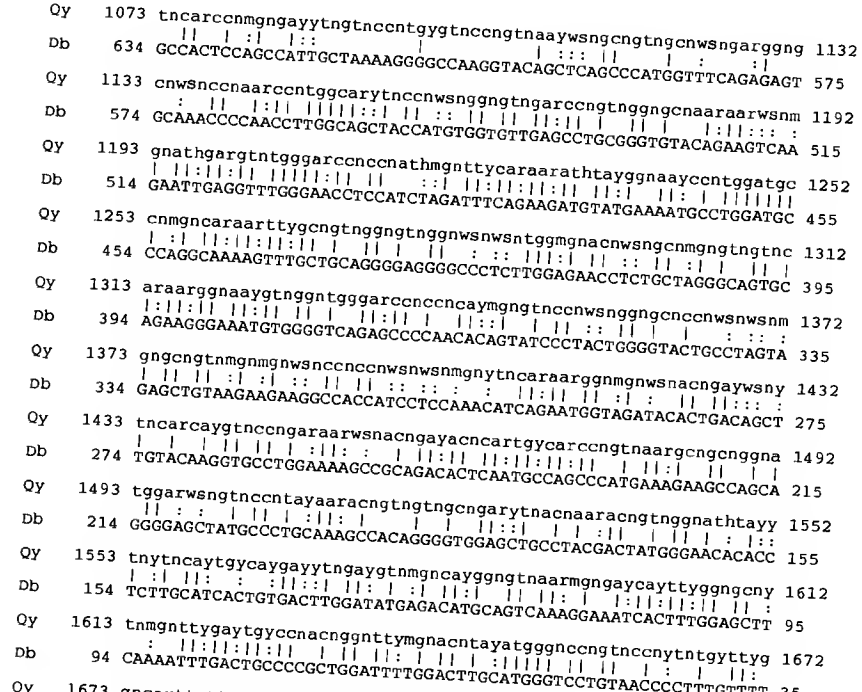
```
Db      420 CATGGAGAACCTCTGCTAGGGCAGTGTGGAAGGGAAATGTGGGGTTGGAGCCCCCGCACA 479

Qy     1346 gngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgny 1405

Db      480 GTGCCCCTGCTGGGGCACTGCCTAGTGGAGCTGTGAGAAGAGGGCCACCATCCTCCAGAC 539

Qy     1406 tncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacnc 1465

Db      540 CCCAGAATGGTAGATCCATCAACAGCTTGCACTGTGCACTTGGAAAAGCCGCAGACACTC 599

Qy     1466 artgycarccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcng 1525

Db      600 GACACCAGCCCATGAAAGCAGCTGGGAGGGAGGCTATACCCTGCAAAGACACAGGGGCAG 659

Qy     1526 arytnacnaaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayg 1585

Db      660 AGCTGCCTAAGACCATGGGAACCCACCTCTTGCATCAGCATGACCTGGATATGAGACTTG 719

Qy     1586 gngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacnt 1645

Db      720 GAGTCAAAGGAGATCATTTTGGAGCTTTGAAATTTGATTGCCCCGCTGGATTTCAGACTT 779

Qy     1646 ayatgggnccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtntty 1704

Db      780 GTGTANGCCTGATAACCCCTTTGTTTTGGGCAATTTCTCCCATTTTGAATGGCTACATTT 839

Qy     1705 acncartgyytntayytncaytgyat 1730

Db      840 TACCAATACCTGTACCCCTATGGGAT 865


RESULT   7
BM457166
LOCUS       BM457166                  898 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6411690 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5583427
            5', mRNA sequence.
ACCESSION   BM457166
VERSION     BM457166.1  GI:18506206
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12346  row: h  column: 20
            High quality sequence start: 9
            High quality sequence stop: 713.
FEATURES    Location/Qualifiers
     source 1. .898
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:5583427"
            /clone_lib="NIH_MGC_92"
            /tissue_type="embryonal carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 2.5 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."

BASE COUNT      225 a    228 c    254 g    191 t
ORIGIN


  Query Match           18.4%;  Score 318.6;  DB 10;  Length 898;
  Best Local Similarity  45.1%;  Pred. No. 1.7e-62;
  Matches  389;  Conservative  162;  Mismatches  253;  Indels   59;  Gaps    5;

Qy      916 tgyathtgyaaygargarytnaaygtnaayccncargayaayggngaraayathwsntgg 975

Db       35 TGCATAAGCAGCAAGGAGCCTAATGTTAATCCCCAAGACCATGGGGAAAATGTCTCCAGG 94

Qy      976 acntgycarmgnwsnwsncarcarwsnathaarwsnytngcntggmgnccnmgnmg---n 1032

Db       95 TCATGTCAGAGACCTTCATGGCAGCCCCTCCCATCACAGGCCCAGAGGCCCAGGAGGAAA 154

Qy     1033 aartggttytgyggnacnggnccnggnwsnytntgytgygtncarccnmgngayytngtn 1092

Db      155 AAGTGGTTTTGTGGGCTGGGCCCAGGGTCTCTGTGCTGTGTGCAGCCTAGGGACTTGGTG 214

Qy     1093 ccntgygtnccng----------------------------------tnaaywsngcngtn 1119

Db      215 CCCTGTGTCCCAGCTGCTCCAGCCATGGCTGAAAGGAGCCAACATAGAGCTTGGGCTGTG 274

Qy     1120 gcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccngtnggn 1179

Db      275 GCTTCAGAGGGTAGAGGCCCCAAGCATTGGCAGCTTCCATGTGATGTTGAGCCTGTGGGT 334

Qy     1180 gcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarathtayggn 1239

Db      335 GCAGAGATGTCAAGAATTGAGATTTGGGAACCTCTGCCTAGATTTCAGAATATATATGGA 394

Qy     1240 aayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsn 1299

Db      395 AACACCTGGATGTCCAGGCAAAAGTTTGCTGCAGGGGCAGAGCCCTCATGGGGAACCTCT 454

Qy     1300 gcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccnwsnggn 1359

Db      455 CCTAGGGCAGTGTGGAAGGGAAATGTGGGG-CAGAGCCCCCACACAGAGTCCCTGTTGGG 513

Qy     1360 gcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnws----------------------n 1398

Db      514 GCACTGCCTAGTGGAGCTGTGAGAAGAAGGCCACAGTCCCAGAACGGTAGACCCCAGAAT 573

Qy     1399 wsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacn 1458

Db      574 GGTAGACCCCAGAATGGTAGATCAACCACCAACTTGCATCTTGTGCCTGGAAAAGCTACA 633

Qy     1459 gayacncartgycarccngtnaargcngcnggnatggarwsngtnccntayaaracngtn 1518

Db      634 GACACTCAACATCAGCCTGAGAAAGCAGCCAGGAGGGAGGCTGTACTCTGCAAAGCCACA 693

Qy     1519 gtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygayytngaygtn 1578

Db      694 GGGGCGGAGCTGCACAAGACCATGGGAACCTATCTCTTGCATCAGCATGACCTGAATGTG 753

Qy     1579 mgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacnggntty 1638

Db      754 AGACATGGAGTCAAAGGAGATCATTATGGAGCTTTAAAATTTGACTGCCCCACTGGATTT 813

Qy     1639 mgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnttyggnacngc 1697

Db      814 CAGACTTGCATGGACCCTGTAACCCCTTTGTTTTGGCCAATTTCTCTCCTTTGGAACAGC 873

Qy     1698 ngtnttyacncartgyytntayy 1720

Db      874 TGTATTTACCCAATACCTGTACC 896


RESULT   8
AG086951
LOCUS       AG086951                  717 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-085H08.F, genomic survey sequence
```

```
ACCESSION   AG086951
VERSION     AG086951.1  GI:16638753
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphobiast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-085H08.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Tayior,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 717)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Tayior,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .717
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-085H08.F"
                     /sex="male"
                     /cell_type="lymphobiast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      175 a    189 c    195 g    158 t
ORIGIN
```

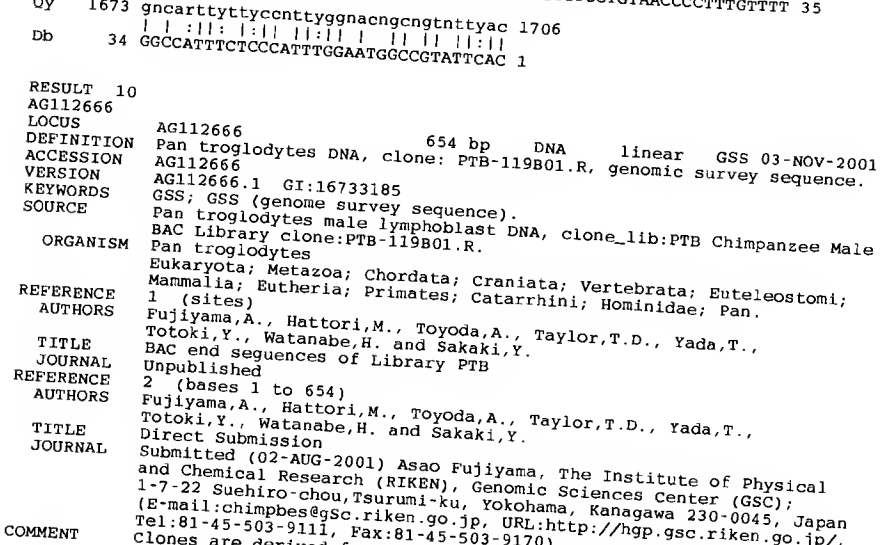
```
  Query Match           18.1%;  Score 313;  DB 12;  Length 717;
  Best Local Similarity 49.1%;  Pred. No. 2.8e-61;
  Matches  299;  Conservative  113;  Mismatches  195;  Indels    2;  Gaps    1;

Qy   1098 ygtnccngtnaaywsngcngtngcnwsngargngcnwsnccnaarccntggcarytncc 1157
          :|| | |     :   || || || :: ||:|| || :: || ||:|| ||||| :: ||
Db     89 CGTCCGCGGAGCTCAGGCTGTGGCTTCAGAGGGTGCAAGCCCTAAGCCTTGGCAGCTTCC 148

Qy   1158 nwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarccnccnat 1217
          || ||  |:|| ||  | ||  |:||::: :| | :| :|| ||||| || |
Db    149 ACATGGTGTTAAGCCTGTGAGTGCACAGAAATCAAGAACTGGGGTTTGGGAACCTCTGCC 208

Qy   1218 hmgnttycaraarathtayggnaaycentggatgccnmgncaraarttygcngtnggngt 1277
          ::| ||:||:| :||:||:|| || : |||||||| :| ||:|| ||:|| || || |
Db    209 TAGATTTCAGAGGATATATGGTAATACCTGGATGCCCAGACAGAATTTTGCTGTAGGGGA 268

Qy   1278 nggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntgggarcc 1337
          ||  : :: || :  || :: || :| | ||   : |: | ||:|| || | ||:||
Db    269 GGGGGCCTCATGAAAAACCTCTGCTAGGGCAGTGTGGGAGAGAAATGTGGGGTCAGAGCC 328

Qy   1338 nccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgngnwsnccnccnws 1397
          | ||::| || || :: || || ||  : ::  | || ||  :  :|  : || |  :
Db    329 ATCACACAGAGTCCCTACTGGGGCACCGCCTAGTGGAGCTGTGAAAAGTGGGCCACTATT 388

Qy   1398 nwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnac 1457
          :: :| :  ||:|| || :| :: || ||::: :| || ||:|  || | :||: : ||
Db    389 CTCCAGACCCCAGAATGGTAGATCCACTGACAGCTTGCACCATGCACCTGGAAAAGCCAC 448

Qy   1458 ngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntayaaracngt 1517
          ||:|| ||:||:||:|| || ||:|| || || | |||::: || || | :||:||
Db    449 AGACACTCAATGCCAGCCTGTGAAAGCAGCTGGGAGGGAGACTGTACCATGCAAGACCAC 508

Qy   1518 ngtngcngarytnacnaaracngtnggnathtayytnytncaytgycaygayytngaygt 1577
          |  || ||:|  | ||:||  | || | : |:|| :| ||:  :||:||::| ||:||
Db    509 AGGGGCAGAGCTGCCCAAGACCATGGGAACCCACCTCTTGCATCATCATGACCTGGATGT 568

Qy   1578 nmgncayggngtnaarmgngaycayttyggngcnytnmgnttygaytgyccnacnggntt 1637
          :| ||:|| |  ||: | ||:||:||:|| || :| :| ||:||:| :||  | || ||
Db    569 GAGACATGGAGCCAAAGGAGATCATTTTGGAGCTTTAAGATTTGACTTCCCTGCTGGATT 628

Qy   1638 ymgnacntayatgggnccngt--nccnytntgyttyggncarttyttyccnttyggnacn 1695
          ::  || | :|||| || ||   || :| | :||:|| ||:||:  :|| ||:|| |
Ob    629 TCAGACTTGCATGGGCCCTGTAGCCCCTTTTTTTTTGGCCAATTTCCCCCATTTGGAATG 688

Qy   1696 gcngtntty 1704
          |  |  |:
Db    689 GGGTTAATT 697
```

```
RESULT   9
BI754555/c                                  736 bp    mRNA    linear   EST 25-SEP-2001
LOCUS       BI754555
DEFINITION  603023384F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5194028 5',
            mRNA sequence.
ACCESSION   BI754555
VERSION     BI754555.1  GI:15746133
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11485  row: g  column: 21
            High quality sequence stop: 733.
FEATURES             Location/Qualifiers
     source          1. .736
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5194028"
                     /clone_lib="NIH_MGC_114"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: EcoRV (destroyed); RNA source anonymous pool of 6
                     male brains, age range 23-27 yo. Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon cloning). Average insert size 1.5 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was constructed by C. Gruber
                     (Invitrogen). Research Genetics tracking code 019. Note:
                     this is a NIH_MGC Library."
BASE COUNT      163 a    196 c    185 g    192 t
ORIGIN
```

```
  Query Match           17.7%;  Score 306.6;  DB 10;  Length 736;
  Best Local Similarity 44.1%;  Pred. No. 8.8e-60;
  Matches  306;  Conservative  122;  Mismatches  266;  Indels    0;  Gaps    0;

Qy   1013 tngcntggmgncngnmgnaartggttytgyggnacnggnccnggnwsnytntgytgyg 1072
          | |  |||:  |   | :  |     | ||: |    || | |   : :| ||:  :
Db    694 TGGTCTGGACCCAAGGTCCCCATGTTGTGTGCAGCCTAGGACTTGCTGCCTTGTGTCCCA 635
```



RESULT 10
AG112666
LOCUS AG112666 654 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-119B01.R, genomic survey sequence.
ACCESSION AG112666
VERSION AG112666.1 GI:16733185
KEYWORDS GSS; GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-119B01.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (sites)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 654)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gSc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers
source 1. .654 .
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-119B01.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 165 a 174 c 180 g 135 t
ORIGIN



RESULT 11
AG064424
LOCUS AG064424 676 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-053H24.R, genomic survey seguence

```
ACCESSION   AG064424
VERSION     AG064424.1  GI:16616226
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-053H24.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 676)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
                Sequencing: M13Rev
            LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES            Location/Oualifiers
     source         1. .676
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /clone="PTB-053H24.R"
                    /sex="male"
                    /cell_type="lymphoblast"
                    /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      164 a    165 c    191 g    156 t
ORIGIN
```

```
Query Match          17.1%;  Score 296;  DB 12;  Length 676;
Best Local Similarity  47.9%;  Pred. No. 2.3e-57;
Matches  294;  Conservative  117;  Mismatches  201;  Indels    2;  Gaps    2;

Qy    1114 gcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 1173
           ||  | || ::  |:|| || :: |  ||:|| ||||::|| ||    |  || ||:|
Db      63 GCCATGGCTTCTAAAGGTGCAAGCCTCAAGCCTTGGCAGCTTCCACATGATGTTGAGCTG 122

Qy    1174 gtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarath 1233
           |  | ||  |:||::: :| | :| :|| | |||:|| |    ::| ||:||:| :||
Db     123 GCAAGTGCACAGAAGTCAAGAACTGGGGTTTAGGAACCTCTACCTAGATTTCAGAGGATG 182

Qy    1234 tayggnaayccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgn 1293
           ||:|| ||: | |||||| | :| ||:||:||:|| |  ||    || :: :: |||:|
Db     183 TATGGAAATGCCTGGATGTCCAGGCAGAAGTTTGCCGCAGGGACGGGGTCCTCTTGGAGA 242

Qy    1294 acnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccn 1353
           || :: || :| |  |  :||:|| ||:|| || | |||: | |  | ::| || ||
Db     243 ACCTCTGCTAGGGCAATGTGGAAGGGAAATGTGGGGTTGGAGTCCCAGCCCAGAGTCCCT 302

Qy    1354 wsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraar 1413
           :: || || || :: ::  | || || :| :| ::  | |  :: :: :| :  ||:||
Db     303 ACTGGGGCACCATCTAGTGGAGCTGTGAGAAG-AGGGCACTGACCTCCAGACCCCAGAAT 361

Qy    1414 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgycar 1473
           || :| :: || ||::: :| || ||:|| || ||:||: :    |:|| ||:  :||:
Db     362 GGTAGATCCACTGACAGCTTACACCATGTGCCTGAAAAAGCTGTAAACACTCAACACCAG 421

Qy    1474 ccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacn 1533
           ||  | || ||  | || | |||·  || || | :| : |    |  || ||::|  |
Db     422 CCCATGAAAGCATCTGGGAGGGAGGTTGTACCCTGCAGATCCACAGGGGCGGAGCTGCCC 481

Qy    1534 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1593
           ||:||  | ||   : |:::| :| ||:     :||::| ||:|| :| ||:|| || ||:
Db     482 AAGACCATGGGAGCCCACCTCTTGCATCAGTGTGACCTGGATGTGAGACATGGAGTCAAA 541

Qy    1594 mgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggn 1653
           | ||:||:||:|  |  :| :| ||:||:||:|   | || ||: | ||   :||||
Db     542 GGAGATCATTTTGTAGTTTTAAGATTTGACTGCCTCGCTGGATTTTGGACTAGCATGGGG 601

Qy    1654 ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg 1712
           || ||  || :| ||:||:|| ||:||: |:|| ||:|| |  || || ||:|| ||:||
Db     602 CCTGTAGCCCCTTTGTTTTGGCCAATTTCTCCCATTTGGAATGGCTGTATTTACCCAATG 661

Qy    1713 yytntayytncayt 1726
           ::| ||::  ||:|
Db     662 CCTGTACCCCCATT 675
```

```
RESULT  12
AG062447                                691 bp    DNA     linear   GSS 03-NOV-2001
LOCUS       AG062447
DEFINITION  Pan troglodytes DNA, clone: PTB-050M20.F, genomic survey sequence.
ACCESSION   AG062447
VERSION     AG062447.1  GI:16614249
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-050M20.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 691)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
                Sequencing: -21M13
            LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES            Location/Oualifiers
     source         1. .691
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /clone="PTB-050M20.F"
                    /sex="male"
                    /cell_type="lymphoblast"
                    /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      160 a    167 c    201 g    162 t      1 others
ORIGIN
```

```
Query Match          16.6%;  Score 287.4;  DB 12;  Length 691;
Best Local Similarity  49.0%;  Pred. No. 2.3e-55;
Matches  301;  Conservative  110;  Mismatches  196;  Indels    7;  Gaps    3;

Qy    1098 ygtnccngtnaaywsngcngtngcnwsngarggngcnwsnccnaarccntggcarytncc 1157
           :|| |  |    :    || || :: || ||:|  |||  ::| ||
Db      82 CGTCCGCGGAGCTCAGGCTGTGGCTTCAGAGGGTGCAAGACCCAAGCTTTGGAGGCTTCC 141
```

```
Qy   1158 nwsnggngtngarccngtnggngcnaaraarwsnmgnathgargtntgggarccnccnat 1217
          :: || || ||:|| ||  | ||  |:||::      ||:| :|| |||||:|| ||  |
Db    142 ATGTGGTGTTGAGCCTGTGAGTGCACAGAAGA-----ATCGGGGTTTGGGAACCTCCACT 196

Qy   1218 hmgnttycaraarathtayggnaayccntggatgccnmgncaraarttygcngtnggngt 1277
          ::| ||:||:||:|| ||:|  ||: | |||||||| :  ||:||:||:|| || ||
Db    197 TAGATTTCAGAAGATGTATGAAAATACCTGGATGCCCATGCAAAAGTTTGCTGTAGGGTC 256

Qy   1278 nggnwsnwsntggmgnacnwsngcnmgngtngtncaraarggnaaygtnggntgggarcc 1337
          ||  : :: |||:| || :: || :| |  ||   :||:|| ||:||  | |  ||: |
Db    257 AGGGCCCTCATGGAGAACTTCTGCTAGGGCAGTGTGGAAGGGAAATGTGAGGTCAGAGGC 316

Qy   1338 nccncaymgngtnccnwsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccncccnws 1397
          || ||::| || || :: || || ||  : ::  | || || :| :|  : || |  ::
Db    317 CCCACACAGAGTCCCTACTGGGGCACCACCCAGTGGAGCTGTGAGAAGAGGGCCACTGTC 376

Qy   1398 nwsnmgnytncaraarggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnac 1457
          :: :| :| ||:|| || :| :: || || :: :| || ||:|| || | :||: : ||
Db    377 CTCCAGACTCCAGAATGGTAGATCCACTGATGGCTTGCATCATGTGCCTGGAAAAGCCAC 436

Qy   1458 ngayacncartgycarccngtnaargcngcnggnatggarwsngtnccntayaaracngt 1517
          ||:|| ||:||:||:||  | ||:|| || || | |||: : || || ||:||: |
Db    437 AGACACTCAATGCCAGCCCATGAAAGCAGCCGGGAAGGAGGCTGTACCCTACAAGGCCAC 496

Qy   1518 ngtngcngaryt-nacnaaracngtnggnathtayytnytncaytgycaygayytngayg 1576
          |  || ||::|  | ||:||  | || | : |::  :| | :  :|||::| | :|
Db    497 AGGGGCAGAGCTGTTCCAAGACCATGGGAACCCACCCCTTGCCTCAGTGTGACCTGGTTG 556

Qy   1577 tnmgncayggngtnaarmgngaycaytttyggngcnytnmgnttygaytgyccnacnggnt 1636
          | :|  |:|| || | : | ||:||:| : | || :| :| ||:||:||:  |  || |
Db    557 TGAGATATGGAGTCAGAGGAGATCATTGTAGAGCTTTAAGATTTGACTGTGTGGCTGGAT 616

Qy   1637 tymgnacntayatgggnccngt-nccnytntgyttyggncarttyttyccnttyggnacn 1695
          |: | || | :|||||  || ||  |  :| ||:||:|| ||:||: |:|| ||:|  |
Db    617 TTGGGACTTGCATGGGGCCTGTAGCTCCTTTGTTTTGGCCAATTTCTCCCATTTGTAATG 676

Qy   1696 gcngtnttyacnca 1709
          || || ||:|| ||
Db    677 GCTGTATTTACCCA 690


RESULT  13
AG116938
LOCUS       AG116938                  720 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-124K01.R, genomic survey sequence.
ACCESSION   AG116938
VERSION     AG116938.1  GI:16737457
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
              BAC Library clone:PTB-124K01.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 720)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
                Sequencing: M13Rev

            LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES                Location/Qualifiers
     source             1. .720
                        /organism="Pan troglodytes"
                        /db_xref="taxon:9598"
                        /clone="PTB-124K01.R"
                        /sex="male"
                        /cell_type="lymphoblast"
                        /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       181 a    186 c    187 g    165 t      1 others
ORIGIN


  Query Match             16.5%;  Score 286.2;  DB 12;  Length 720;
  Best Local Similarity   48.5%;  Pred. No. 4.5e-55;
  Matches  298;  Conservative  112;  Mismatches  200;  Indels    4;  Gaps    3;

Qy   1114 gcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 1173
           ||  | || :: || || || ::  | ||: | |||  ::| ||    ||  | ||:||
Db    110 GCCATGGCTTCAGATGGTGCAAGCACCAAGTCTTGGTGGCTTCCACATGGCATTGAGCCT 169

Qy   1174 gtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarath 1233
           |  || ||  |:||::: :| ||:||:|| ||| |:|| ||   :  ||:||:|| ||
Db    170 GCGGGTGCACAGAAGTCAAGAATTGAGGTTTGGAAACCTCCACCTGAATTTCAGAACATG 229

Qy   1234 tayggnaaycentggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgn 1293
          ||:|| ||: | |||||||| :| ||:||:||:|| |  || |   |    :: ||:||
Db    230 TATGGAAATGCCTGGATGCCCAGGCAGAAGTTTGCTGCAGGGGAGTGGCTCTCATGGAGA 289

Qy   1294 acnwsngcnmgngtngtncaraarggnaaygtnggntgggarccnccncaymgngtnccn 1353
          || :: || :| |  || ||:||:|| ||:|| || | |||:|| || ||::| || ||
Db    290 ACCTCTGCTAGGGCAGTGCAGAAGGGAAATGTGGGGTCGGAGCCCCCACACAGGGTCCCT 349

Qy   1354 wsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraar 1413
          :: || |   |  : ::  | || || :| :|   || |  :: :: :| :  ||:||
Db    350 ACTGGGG-ATCACCTAGTGGAGCTGTGAGAAGACAGCCACTGTCTTCCAGACCCCAGAAT 408

Qy   1414 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgycar 1473
          |  :| :: || ||::: :   :| :|| || | :||: :  | | :|| ||:  :||:
Db    409 GCTAGATCCACTGACACTTGC--ACCTGTGCCTGGAAAAGCTGCAGGTACTCAACATCAG 466

Qy   1474 ccngtnaargcngcnggngatggarwsngtnccntayaaracngtngtngcngarytnacn 1533
          ||  | ||:|| || || | |||: : || || | :||: |   |  |  ||::|  |
Db    467 CCCATGAAAGCAGCTGGGAGGGAGGCTGTACCCTGCAAAGCAACAGAGGTGGAGCTTCCC 526

Qy   1534 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1593
          |:||  | || | :||::| :| ||:     :|||::| ||:|| :| ||:||  | ||:
Db    527 GAGACCATGGGAACCTACCTCTTGCATCAGGGTGACCTGGATGTGAGACATGGATTCAAA 586

Qy   1594 mgngaycaytttyggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggn 1653
          :| ||:||: |:|| || :| :| ||:||: |:|  || || ||: | || | :|||||
Db    587 AGAGATCATATTGGAGCTTTAAGATTTGACAGCCACACTGGATTTTGGACTTGCATGGGC 646

Qy   1654 ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg 1712
          ||  ||  || :| ||:||:|| ||:||: |:|| ||:|| |  |  |  ||:||  |:||
Db    647 CCTGTAGCCCTTTTGTTTTGGCCAATTTCTCCCATTTGGAATGGTTGCATTTACCTAATG 706

Qy   1713 yytntayytncayt 1726
          ::| ||::  ||:|
Db    707 TCTGTATTCCCATT 720


RESULT  14
AQ037711
LOCUS       AQ037711                  623 bp    DNA     linear   GSS 11-JUL-1998
DEFINITION  CIT-HSP-2337G10.TR CIT-HSP Homo sapiens genomic clone 2337G10, DNA
            sequence.
ACCESSION   AQ037711
VERSION     AQ037711.1  GI:3303543
KEYWORDS    GSS.
```

```
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2337G10.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .623
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2337G10"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      160 a    146 c    172 g    145 t
ORIGIN
```

```
  Query Match             16.4%;  Score 283.8;  DB 12;  Length 623;
  Best Local Similarity   48.1%;  Pred. No. 1.5e-54;
  Matches  286;  Conservative  111;  Mismatches  196;  Indels    2;  Gaps    2;

Qy   1138 ccnaarccntggcarytnccnwsnggngtngarccngtnggngcnaaraarwsnmgnath 1197
          |  ||:|  ||||::|  ||    || || ||: | |  ||  |   :||::: :| | :
Db     26 CTCAAGCTTTGGCAGCTTCCACATGGTGTTGAGTCTGCGGGTACACCAAAGTCAAGAACT 85

Qy   1198 gargtntg-ggarccnccnathmgnttycaraarathtayggnaaycentggatgccnmg 1256
          ||:|| || |||:|| ||   ::  ||:||:||:|| ||:|| ||: | |||||| | :|
Db     86 GAGGTTTGAGGAACCTCCACCTACATTTCAGAAGATGTATGGAAATGCCTGGATGTCCAG 145

Qy   1257 ncaraarttygcngtnggngtnggnwsnwsntggmgnacnwsngcnmgngtngtncaraa 1316
           ||:||:||:|| |  || |  ||  : :: |||:| || :: || :| |  || | :||
Db    146 GCAAAAGTTTGCTGCAGGGGCAGGGCCCAGATGGAGAACCTCTGCTAGGGCAGTGCGGAA 205

Qy   1317 rggnaaygtnggntgggarccnccncaymgngtnccnwsnggngcnccnwsnwsnmgngc 1376
          :|| ||:|| || | |||:||  | || :| ||  | :: || || |   : ::  | ||
Db    206 GGGAAATGTGGGGTTGGAGCCGTCACAGAGAGTCTCTACTGGGGCACTGCCTAGTGGAGC 265

Qy   1377 ngtnmgnmgnwsnccnccnwsnwsnmgnytncaraarggnmgnwsnacngaywsnytnca 1436
           || :| :|  : ||    :: :: :| :   |:|| || :| ::  | ||::: :| ||
Db    266 TGTGAGAAGAGGGCCATTGTCCTCCAGACCCTAGAATGGTAGATCTGCCGACAGCTTGCA 325

Qy   1437 rcaygtnccngaraarwsnacngayacncartgycarccngtnaargcngcnggnatgga 1496
            |:|   | | :||: : || ||:|| ||:  :||:||  | ||: | ||  | | ||
Db    326 TTATGCAACTGGAAAAGCCACAGACACTCAACACCAGCCCATGAAAACAGCCAGGAGGGG 385

Qy   1497 rwsngtnccntayaaracngtngtngcngarytnacnaaracngtnggnathtayytnyt 1556
          : :  | || | :||: |    |  || ||::   |  |: |  | || | : |::| :|
Db    386 AGCCATACCCTGCAAAGCCACAGGGGCAGAGTGGCCTGAGGCCATGGGAACCCACCTCTT 445

Qy   1557 ncaytgycaygayytngaygtnmgncayggngtnaarmgngaycayttyggngcnytnmg 1616
           ||:  | :||::  ||:|| :| ||:|| || ||: | ||:||:||:|| || :| :|
Db    446 GCATCAGCGTGACCCAGATGTGAGACATGGTGTCAAAGGAGATCATTTTGGAGCTTTAAG 505
```

```
Qy   1617 nttygaytgyccnacnggnttymgnacntayatgggnccngt-nccnytntgyttyggnc 1675
           ||:||:||:||  | || ||::  || | :| ||| || ||  || :| | :||:|| |
Db    506 ATTTGACTGCCCTGCTGGATTTCAGACTTGTAAGGGGCCAGTAGCCCTTTTCTTTTGGCC 565

Qy   1676 arttyttyccnttyggnacngcngtnttyacncartgyytntayytncaytgyat 1730
          |:||: |:|| ||:|| |  || || ||:|| ||:||::| ||::  ||:||:||
Db    566 AATTTCTCCCATTTGGAATGGCTGTATTTACCCAATGCCTGTATTCCCATTGTAT 620


RESULT  15
AG097258
LOCUS       AG097258                677 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-098001.R, genomic survey sequence.
ACCESSION   AG097258
VERSION     AG097258.1  GI:16717775
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-098001.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 677)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .677
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-098001.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      178 a    175 c    176 g    147 t      1 others
ORIGIN
```

```
  Query Match             16.4%;  Score 283.2;  DB 12;  Length 677;
  Best Local Similarity   48.8%;  Pred. No. 2.1e-54;
  Matches  294;  Conservative  113;  Mismatches  187;  Indels    9;  Gaps    3;

Qy   1114 gcngtngcnwsngarggngcnwsnccnaarccntggcarytnccnwsnggngtngarccn 1173
          || || || :: ||:||    :: || ||:|| |||||::| || :: || || ||:||
Db     68 GCTGTGGCTTCAGAAGGTAGAAGCCCCAAGCCTTGGCAGCTTCCATGTGGTGTTGAGCCT 127

Qy   1174 gtnggngcnaaraarwsnmgnathgargtntgggarccnccnathmgnttycaraarath 1233
          || ||  |:||::: :  ||:||:|| || ||:|| ||   ::| ||: |:| :||
Db    128 ACAGGTGCACAGAAGTCACAAATTGAGGTTTGTGAACCTCCACCTAGATTT-AGACGATG 186

Qy   1234 tayggnaaygccntggatgccnmgncaraarttygcngtnggngtnggnwsnwsntggmgn 1293
          ||:|| ||: | |||||||| :|  |:||:||:|| |  ||     |  : :: |||:|
Db    187 TATGGAAACACCTGGATGCCCAGGTAAAAGTTTGCTGCAGGGTGGAGGCCATCATGGAGA 246
```

```
Qy   1294 acnwsngcnmgngtngtncaraarggnaaygtnggntgggarccncccncaymgngtnccn 1353
          || ::  | :| |  || | :||:|| ||:|  || | |||:|| ||  |::| || ||
Db    247 ACCTCTTCTAGGGCAGTACGGAAAGGAAATGCAGGGTTGGAGCCCCCAAACAGAGTCCCT 306

Qy   1354 wsnggngcnccnwsnwsnmgngcngtnmgnmgnwsnccnccnwsnwsnmgnytncaraar 1413
          :  || || |   : ::    || || :| :|  : || || :: :: :| :  ||:||
Db    307 AATGGCGCACTGCCTAGTGAAGCTGTAAGAAGAGGACCACCATCCTCCAGACCCCAGAAT 366

Qy   1414 ggnmgnwsnacngaywsnytncarcaygtnccngaraarwsnacngayacncartgycar 1473
          || :| :: ||  |::: :| ||:||:|  || | :||: :  | ||:|| ||: |:||:
Db    367 GGTAGATCCACCTACAGCTTGCAACATGCACCTGGAAAAGCTGCAGACACTCAGCGCCAG 426

Qy   1474 ccngtnaargcngcnggnatggarwsngtnccntayaaracngtngtngcngarytnacn 1533
          ||  | ||:|| ||  | | |||::: ||  | | :||: |    |  || ||::|  |
Db    427 CCCATGAAAGCAGCCAGGAGGGAGACTGTCTCCTGCAAATCCACAGGGGCAGAGCTGCCC 486

Qy   1534 aaracngtnggnathtayytnytncaytgycaygayytngaygtnmgncayggngtnaar 1593
          ||:||  | || | : |::| :| ||:   | :||::| ||:|| :| ||:|| || ||:
Db    487 AAGACCATGGGAACCCACCTCTTGCATCAACGTGACCTGGATGTGAGGCATGGAGTCAAA 546

Qy   1594 mgngaycayttyggngcnytnmgnttygaytgyccnacnggnttymgnacntayatgggn 1653
           | ||:||:| :|| ||       ||:||:||:||  | || ||: | || | :|||||
Db    547 GGAGATCATTNTGGAGC-------TTTGATTGCCCTGCTGGATTTTGGACTTGCATGGGC 599

Qy   1654 ccngt-nccnytntgyttyggncarttyttyccnttyggnacngcngtnttyacncartg 1712
          || ||  || :| | :||:|  ||:||: |:|| ||:|  |  || || ||:|| ||:||
Db    600 CCTGTATCCCCTGTATTTTGAGCAATTTATCCCATTTGAAATGGCTGTATTTACCCAATG 659

Qy   1713 yyt 1715
          ::|
Db    660 CCT 662
```

Search completed: July 8, 2002, 18:51:07
Job time: 19347 sec